GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 04:03:35 ; Search time 303 Seconds
(without alignments)
4072.921 Million cell updates/sec

Title: US-09-744-679-7
Perfect score: 548
Sequence: 1 cgccgagggtcgcttggacc.....ccatcgttgtgttcatcctg 548

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	548	100.0	548	21	AAZ50772	Human thrombin rec
2	548	100.0	3472	20	AAZ32191	Human thrombin rec
3	548	100.0	3480	13	AAQ28568	Human thrombin rec
4	548	100.0	3480	21	AAZ50771	Human thrombin rec
5	387	70.6	1278	24	ABK70889	Human cDNA encodin
6	328.4	59.9	2910	18	AAT62461	ThR-GPA1 fusion ge
7	322.4	58.8	6546	21	AAF21433	Human low adenosin
8	322.4	58.8	6546	21	AAA35311	Human adenosine re
9	321	58.6	3182	21	AAF21432	Human low adenosin
10	321	58.6	3182	21	AAA35310	Human adenosine re
11	282.4	51.5	1764	15	AAQ73590	Fragment of the hu
12	249	45.4	772	23	AAS68315	DNA encoding novel
13	240.8	43.9	1209	24	ABK70887	Human cDNA encodin
14	184.2	33.6	1312	18	AAT93367	Mouse thrombin rec
c 15	148.2	27.0	6436	24	ABL32681	Human immune syste
16	147	26.8	1116	24	ABK70888	Human cDNA encodin
17	146.6	26.8	6436	24	ABL32680	Human immune syste
18	146	26.6	572	21	AAA27044	Human cell surface
c 19	137.2	25.0	556	24	ABQ23658	Oligonucleotide fo
20	137.2	25.0	556	24	ABQ23659	Oligonucleotide fo
21	135.6	24.7	556	24	ABQ23660	Oligonucleotide fo
c 22	135.6	24.7	556	24	ABQ23661	Oligonucleotide fo
23	126.6	23.1	1300	19	AAV33372	PROLAC FLAG-connec
24	125.4	22.9	1300	19	AAV33371	PROLAC FLAG-connec
c 25	56	10.2	56	19	AAV33375	Fragment of thromb
26	55	10.0	55	19	AAV33376	Fragment of thromb
27	48.8	8.9	47981	22	AAF30757	Micromonospora meg
28	46	8.4	54	19	AAV33374	Fragment of thromb
29	44.6	8.1	192	18	AAT97178	Thrombin site-DR b
30	44.4	8.1	12111	22	ABA21422	Human nervous syst
c 31	42	7.7	3272	24	ABI99475	Mouse ischaemic co
32	41.8	7.6	2259	23	AAS79424	DNA encoding novel
33	40.8	7.4	67212	21	AAA08954	WFS1 variant genom
34	40.6	7.4	189	18	AAT97177	Thrombin site-DR a
c 35	40.6	7.4	2341	19	AAV62155	HSV-2 strain SB5 C
36	40.6	7.4	2694	19	AAV62145	HSV-2 strain SB5 C
c 37	40.6	7.4	117213	19	AAV62176	HSV-2 strain SB5 C
c 38	40.6	7.4	154746	24	AAD25519	Human herpesvirus
39	40.4	7.4	327	23	AAH88285	CNS disorder-relat
40	40.2	7.3	32195	22	AAS36105	Human cardiovascul
41	40.2	7.3	32195	22	AAS31532	Human DNA for a no
42	40.2	7.3	32195	22	AAS31538	Human DNA for a no
43	40.2	7.3	32195	23	ABK44045	Genomic DNA encodi
44	40.2	7.3	32195	24	ABQ66856	Human polynucleoti
45	40.2	7.3	32195	24	ABQ66862	Human polynucleoti

ALIGNMENTS

RESULT 1
AAZ50772
ID AAZ50772 standard; cDNA; 548 BP.
XX
AC AAZ50772;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human thrombin receptor antisense cDNA.
XX
KW Human; thrombin receptor; ThR; PAR-1; protease activated receptor;
KW antisense molecule; PAR antibody; cytostatic; therapeutic;
KW metastatic tumour cell; placental implantation; invasive cell; ss.
XX
OS Homo sapiens.
XX
PN WO200008150-A1.
XX
PD 17-FEB-2000.
XX
PF 05-FEB-1999; 99WO-IL00079.
XX
PR 07-AUG-1998; 98IL-0125698.
XX
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
PI Bar-Shavit R;
XX
DR WPI; 2000-205706/18.
XX
PT Treating metastatic tumor cells useful for treating disorders involving
PT placenta implantation in a female comprises administration of an

PT antisense molecule complementary to an RNA sequence of a protease
PT activated receptor protein -
XX
PS Claim 6; Fig 2; 46pp; English.
XX
CC The patent discloses a method to treat metastatic tumour cells using
CC an antisense molecule comprising a polynucleotide complementary to an
CC RNA sequence of a protease activated receptor (PAR) protein, or an
CC antibody capable of binding to a PAR protein. The antisense molecules and
CC antibodies of PAR protein are also used to treat disorders associated
CC with implantation of placenta. The present sequence is an antisense
CC cDNA of human thrombin receptor (ThR) (also known as PAR-1), which is a
CC G-coupled protein belonging to the PAR family. This sequence is useful
CC in the treatment of invasive cells.
XX
SQ Sequence 548 BP; 122 A; 166 C; 148 G; 112 T; 0 other;

Query Match 100.0%; Score 548; DB 21; Length 548;
Best Local Similarity 100.0%; Pred. No. 6.8e-139;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
Db       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
Db      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
Db     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
Db     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
Db     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
Db     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360

Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
Db     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
Db     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
Db     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540

Qy     541 TCATCCTG 548
Db     541 TCATCCTG 548
```

RESULT 2
AAZ32191
ID AAZ32191 standard; cDNA; 3472 BP.
XX
AC AAZ32191;
XX
DT 13-JAN-2000 (first entry)
XX
DE Human thrombin receptor nucleotide sequence.
XX
KW Human; coding sequence polymorphism; vascular pathology gene;
KW polymorphic site; phenotype correlation; forensic; paternity testing;
KW medicine; genetic analysis; vascular disease; ds.
XX
OS Homo sapiens.
XX
PN WO9950454-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-US06473.
XX
PR 01-APR-1998; 98US-0054272.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Lander ES, Daley GQ, Cargill M, Ireland JS, Rozen SG;
XX
DR WPI; 1999-620066/53.
DR P-PSDB; AAY49570.
XX
PT Determination of polymorphisms in genes, especially those identifying
PT predisposition to vascular disease -
XX
PS Claim 1; Fig 35; 134pp; English.
XX
CC AAZ32159 to AAZ32194 represent reference alleles for specifically
CC claimed nucleic acid sequences from the present invention which comprise
CC polymorphic sites as given in a table in the specification, selected
CC from 92 single nucleotide polymorphisms in which the nucleotide at the
CC polymorphic site is different from a nucleotide at the same site in a
CC reference allele. The nucleic acids, and primers and probes, are used to
CC identify polymorphisms, which may predispose an individual to disease,
CC especially a vascular disease. They can also be used in phenotype
CC correlations, forensics, paternity testing, medicine or genetic
CC analysis. AAY49550 to AAY49573 represent the proteins which correspond
CC to some of the reference alleles.
XX
SQ Sequence 3472 BP; 933 A; 817 C; 785 G; 937 T; 0 other;

Query Match 100.0%; Score 548; DB 20; Length 3472;
Best Local Similarity 100.0%; Pred. No. 1.4e-138;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
Db      64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
Db     124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
Db     184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
Db     244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
Db     304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
Db     364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
Db     424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
```

```
Db     484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603

Qy     541 TCATCCTG 548
           ||||||||
Db     604 TCATCCTG 611
```

RESULT 3
AAQ28568
ID AAQ28568 standard; DNA; 3480 BP.
XX
AC AAQ28568;
XX
DT 15-FEB-1993 (first entry)
XX
DE Human thrombin receptor gene.
XX
KW Diagnosis; cardiovascular disease; wound healing; restenosis;
KW thrombosis; unstable angina treatment; myocardial infarction;
KW thrombotic; thromboembolytic stroke; ss.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	225..1502
FT		/*tag= a
FT	sig_peptide	225..303
FT		/*tag= b
FT	mat_peptide	304..1502
FT		/*tag= c

XX
PN WO9214750-A.
XX
PD 03-SEP-1992.
XX
PF 19-FEB-1992; 92WO-US01312.
XX
PR 19-FEB-1991; 91US-0657769.
PR 07-NOV-1991; 91US-0789184.
XX
PA (CORT-) COR THERAPEUTICS INC.
PA (REGC) UNIV CALIFORNIA.
XX
PI Coughlin SR, Scarborough RM;
XX
DR WPI; 1992-316119/38.
DR P-PSDB; AAR27240.
XX
PT DNA encoding cell surface receptor for thrombin - useful for
PT determining thrombin in diagnosing e.g. cardiovascular diseases,
PT also to treat wound healing, restenosis etc.
XX
PS Disclosure; Fig 1; 81pp; English.
XX
CC The sequence is that of the gene coding for human thrombin receptor.
CC It can be used in the prepn. of diagnostics to determine thrombin
CC levels in samples, and screening tools for candidate substances which
CC affect thrombin activity in vivo. Thrombosis may be diagnosed in a
CC mammal by measuring the presence, absence or amt. of the cleaved
CC activation peptide of the TR.
XX
SQ Sequence 3480 BP; 947 A; 816 C; 786 G; 931 T; 0 other;

Query Match 100.0%; Score 548; DB 13; Length 3480;
Best Local Similarity 100.0%; Pred. No. 1.4e-138;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603

Qy     541 TCATCCTG 548
           ||||||||
Db     604 TCATCCTG 611
```

RESULT 4
AAZ50771
ID AAZ50771 standard; DNA; 3480 BP.
XX
AC AAZ50771;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human thrombin receptor DNA.
XX
KW Human; thrombin receptor; ThR; PAR-1; protease activated receptor;
KW antisense molecule; PAR antibody; cytostatic; therapeutic;
KW metastatic tumour cell; placental implantation; invasive cell; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	225..1502
FT		/*tag= a
FT		/product= "Thrombin receptor"
FT		/transl_except= (pos:1500..1502, aa:Glx)
FT		/note= "no stop codon given"
FT	misc_binding	320..570
FT		/*tag= b
FT		/bound_moiety= "ThR RNA probe"

XX
PN WO200008150-A1.
XX
PD 17-FEB-2000.
XX
PF 05-FEB-1999; 99WO-IL00079.
XX
PR 07-AUG-1998; 98IL-0125698.
XX

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
PI Bar-Shavit R;
XX
DR WPI; 2000-205706/18.
DR P-PSDB; AAY45035.
XX
PT Treating metastatic tumor cells useful for treating disorders involving
PT placenta implantation in a female comprises administration of an
PT antisense molecule complementary to an RNA sequence of a protease
PT activated receptor protein -
XX
PS Example 2; Fig 1a; 46pp; English.
XX
CC The patent discloses a method to treat metastatic tumour cells using
CC an antisense molecule comprising a polynucleotide complementary to an
CC RNA sequence of a protease activated receptor (PAR) protein, or an
CC antibody capable of binding to a PAR protein. The antisense molecules and
CC antibodies of PAR protein are also used to treat disorders associated
CC with implantation of placenta. The present sequence is a DNA encoding
CC human thrombin receptor (ThR) (also known as PAR-1), which is a
CC G-coupled protein belonging to the PAR family. This sequence is used to
CC produce antisense molecules or antibodies against PAR protein for
CC therapeutic use.
XX
SQ Sequence 3480 BP; 941 A; 817 C; 785 G; 937 T; 0 other;

Query Match 100.0%; Score 548; DB 21; Length 3480;
Best Local Similarity 100.0%; Pred. No. 1.4e-138;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
Db      64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
Db     124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
Db     184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
Db     244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
Db     304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
Db     364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
Db     424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
Db     484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
Db     544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603

Qy     541 TCATCCTG 548
Db     604 TCATCCTG 611
```

RESULT 5

ABK70889
ID ABK70889 standard; cDNA; 1278 BP.
XX
AC ABK70889;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human cDNA encoding PAR1 type thrombin receptor.
XX
KW Human; ss; gene; PAR1; thrombin receptor; antiinflammatory; cytostatic;
KW inflammatory disease; cell proliferative disease.
XX
OS Homo sapiens.
XX
PN JP2002010784-A.
XX
PD 15-JAN-2002.
XX
PF 29-JUN-2000; 2000JP-0196514.
XX
PR 29-JUN-2000; 2000JP-0196514.
XX
PA (TEIJ) TEIJIN LTD.
XX
DR WPI; 2002-321520/36.
DR P-PSDB; ABG35300.
XX
PT An inhibitor of cell growth mediated by thrombin used to treat
PT inflammatory and cell proliferative diseases -
XX
PS Claim 9; Page 26-27; 44pp; Japanese.
XX
CC The invention relates to a polypeptide or a compound which can inhibit
CC cell growth caused by thrombin. The polypeptide/compound combines to a
CC specific region of the structure of PAR1 type human thrombin receptor
CC participating to cell growth. Preferably, the compound contains the
CC 52nd to the 56th amino acid sequences at the amino end side of PAR1 type
CC human thrombin receptor ((X4)-Tyr-Glu-Pro-Phe-Trp-(X5) X4, X5 = optional
CC amino acid or peptide sequence). Also included are a modified PAR1 type
CC thrombin receptor gene or its fragment used for obtaining the above
CC polypeptide, a human PAR1 type thrombin receptor protein and its
CC encoding DNA comprising a fully. The polypeptide or the compound is used
CC to treat inflammatory diseases and cell proliferative diseases. The
CC present sequence is a cDNA encoding a PAR1 type thrombin receptor (or a
CC modified version).
XX
SQ Sequence 1278 BP; 276 A; 353 C; 284 G; 365 T; 0 other;

Query Match 70.6%; Score 387; DB 24; Length 1278;
Best Local Similarity 100.0%; Pred. No. 4.8e-95;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     162 ATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTG 221
Db       1 ATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTG 60

Qy     222 TCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCC 281
Db      61 TCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCC 120

Qy     282 CGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAG 341
Db     121 CGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAG 180

Qy     342 AAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTT 401
Db     181 AAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTT 240

Qy     402 CAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGG 461
Db     241 CAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGG 300

Qy     462 CTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAAC 521
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAAC 360

Qy     522 ATCATGGCCATCGTTGTGTTCATCCTG 548
           ||||||||||||||||||||||||||||
Db     361 ATCATGGCCATCGTTGTGTTCATCCTG 387


RESULT 6
AAT62461
ID   AAT62461 standard; DNA; 2910 BP.
XX
AC   AAT62461;
XX
DT   16-AUG-1997  (first entry)
XX
DE   ThR-GPA1 fusion gene.
XX
KW   G-protein coupled receptor; agonist; antagonist; assay;
KW   G-alpha protein; Gpa1p; GPA1 gene; thrombin receptor; ds.
XX
OS   Chimaeric Homo sapiens;
OS   Chimaeric Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2679
FT                   /*tag=  a
FT                   /product= ThR-GPA1 fusion
FT   mRNA            1..39
FT                   /*tag=  b
FT                   /note= "STET nucleotides 535-573"
FT   mRNA            40..1251
FT                   /*tag=  c
FT                   /product= thrombin receptor
FT                   /note= "corresponds to nucleotides 288-1499 of
FT                         the thrombin receptor gene"
FT   mRNA            1252..1260
FT                   /*tag=  d
FT                   /note= "STET nucleotides 1827..1835"
FT   mRNA            1261..2910
FT                   /*tag=  e
FT                   /product= Gpa1p
FT                   /note= "corresponds to nucleotides 202-1851 of
FT                         GPA1"
FT   primer_bind     complement (42..62)
FT                   /*tag=  f
FT                   /note= "oligonucleotide g for thrombin receptor
FT                         gene amplification"
FT   primer_bind     1229..1251
FT                   /*tag=  g
FT                   /note= "oligonucleotide h for thrombin receptor
FT                         gene amplification"
XX
PN   WO9711159-A1.
XX
PD   27-MAR-1997.
XX
PF   20-SEP-1996;   96WO-US15203.
XX
PR   20-SEP-1995;   95US-0004023.
XX
PA   (HEAR-) HEARTLAND BIOTECHNOLOGIES LLC.
XX
PI   Boulton TG,  Das PO,  Mandell RB,  McMullen TW;
XX
DR   WPI; 1997-202868/18.
DR   P-PSDB; AAW16314.
XX
PT   New yeast constructs for use in screening assays - which express a
PT   fusion comprising a mammalian or fungal transmembrane receptor and
PT   a yeast G-alpha protein
XX
PS   Example 2; Fig 12A-G; 109pp; English.
XX
CC   A gene fusion (AAT62461) codes for a polypeptide (AAW16314) comprising
CC   the human thrombin receptor (ThR) covalently linked to yeast G-alpha
CC   protein Gpa1p (encoded by the GPA1 gene).  It was obtd. by PCR
CC   amplification (see also AAT62466-67) of the ThR gene and insertion
CC   of the amplified sequence into pRMHBT3, which contains the GPA1
CC   gene.  Thrombin-dependent activation of the yeast mating pathway
CC   by the ThR-GPA1 fusion protein was observed in ste2 gpa1- yeast
CC   cells.  The polypeptide is an example of novel fusion proteins
CC   between eukaryotic heterotrimeric G-protein coupled receptors and
CC   yeast G-alpha proteins that can be coupled to the pheromone-
CC   induced signal transduction pathway of yeast and used in screening
CC   assays to identify agonists or antagonists of the receptor.
XX
SQ   Sequence 2910 BP; 868 A; 606 C; 628 G; 808 T; 0 other;

  Query Match             59.9%;  Score 328.4;  DB 18;  Length 2910;
  Best Local Similarity   98.2%;  Pred. No. 5.3e-79;
  Matches  332;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy     211 GCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCA 270
           ||   || || | ||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 GCAATCTATTTTATGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCA 85

Qy     271 CCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGG 330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86 CCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGG 145

Qy     331 AGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAA 390
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     146 AGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAA 205

Qy     391 GCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGA 450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206 GCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGA 265

Qy     451 CCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCC 510
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     266 CCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCC 325

Qy     511 TCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
           ||||||||||||||||||||||||||||||||||||||
Db     326 TCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 363


RESULT 7
AAF21433
ID   AAF21433 standard; DNA; 6546 BP.
XX
AC   AAF21433;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #3000.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
```

XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1416-1418; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 6546 BP; 1603 A; 1691 C; 1500 G; 1751 T; 1 other;

Query Match 58.8%; Score 322.4; DB 21; Length 6546;
Best Local Similarity 99.7%; Pred. No. 3.1e-77;
Matches 323; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    225 GCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGG 284
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3362 GCACGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGG 3421

Qy    285 TCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAA 344
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3422 TCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAA 3481

Qy    345 AATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAA 404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3482 AATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAA 3541

Qy    405 AAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTG 464
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3542 AAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTG 3601

Qy    465 ACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATC 524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3602 ACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATC 3661

Qy    525 ATGGCCATCGTTGTGTTCATCCTG 548
```

```
          ||||||||||||||||||||||||
Db   3662 ATGGCCATCGTTGTGTTCATCCTG 3685
```

RESULT 8
AAA35311
ID AAA35311 standard; DNA; 6546 BP.
XX
AC AAA35311;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:185.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 1331-1333; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 6546 BP; 1603 A; 1692 C; 1500 G; 1750 T; 1 other;

```
  Query Match           58.8%;  Score 322.4;  DB 21;  Length 6546;
  Best Local Similarity 99.7%;  Pred. No. 3.1e-77;
  Matches  323;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     225 GCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGG 284
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3362 GCACGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGG 3421

Qy     285 TCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAA 344
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3422 TCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAA 3481

Qy     345 AATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAA 404
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3482 AATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAA 3541

Qy     405 AAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTG 464
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3542 AAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTG 3601

Qy     465 ACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATC 524
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3602 ACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATC 3661

Qy     525 ATGGCCATCGTTGTGTTCATCCTG 548
           ||||||||||||||||||||||||
Db    3662 ATGGCCATCGTTGTGTTCATCCTG 3685
```

```
RESULT 9
AAF21432
ID   AAF21432 standard; DNA; 3182 BP.
XX
AC   AAF21432;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2999.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1415-1416; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 3182 BP; 895 A; 701 C; 684 G; 902 T; 0 other;
```

```
  Query Match           58.6%;  Score 321;  DB 21;  Length 3182;
  Best Local Similarity 100.0%;  Pred. No. 5.6e-77;
  Matches  321;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     228 CGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCA 287
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCA 60

Qy     288 TTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAAT 347
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAAT 120

Qy     348 GAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAA 407
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAA 180

Qy     408 CAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACA 467
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACA 240

Qy     468 CTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATG 527
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATG 300

Qy     528 GCCATCGTTGTGTTCATCCTG 548
           |||||||||||||||||||||
Db     301 GCCATCGTTGTGTTCATCCTG 321
```

```
RESULT 10
AAA35310
ID   AAA35310 standard; DNA; 3182 BP.
XX
AC   AAA35310;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide 2nd SEQ ID NO:184.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
```

KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 1330-1331; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 3182 BP; 895 A; 701 C; 684 G; 902 T; 0 other;

Query Match 58.6%; Score 321; DB 21; Length 3182;
Best Local Similarity 100.0%; Pred. No. 5.6e-77;
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    228 CGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCA 287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCA 60

Qy    288 TTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAAT 347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAAT 120

Qy    348 GAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAA 407
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAA 180

Qy    408 CAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACA 467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACA 240

Qy    468 CTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATG 527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATG 300

Qy    528 GCCATCGTTGTGTTCATCCTG 548
          |||||||||||||||||||||
Db    301 GCCATCGTTGTGTTCATCCTG 321
```

RESULT 11
AAQ73590
ID AAQ73590 standard; DNA; 1764 BP.
XX
AC AAQ73590;
XX
DT 25-JUN-1995 (first entry)
XX
DE Fragment of the human thrombin receptor gene.
XX
KW TR; expression; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 376..1653
FT /*tag= a
XX
PN WO9421789-A.
XX
PD 29-SEP-1994.
XX
PF 28-FEB-1994; 94WO-US02388.
XX
PR 25-MAR-1993; 93US-0038662.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Betlach MC, Turner GJ;
XX
DR WPI; 1994-317010/39.
DR P-PSDB; AAR60698.
XX
PT Expression of heterologous proteins in halo-bacteria - using
PT regulatory and stop sequences from halo-bacteria, pref. the
PT bacterio-rhodopsin gene.
XX
PS Disclosure; Fig 14; 118pp; English.
XX
CC The sequence is that of the human thrombin receptor gene fragment.
CC The gene is used to exemplify a new expression vector for producing
CC heterologous polypeptides in a halobacterial host.
CC See also AAQ73586-92.
XX
SQ Sequence 1764 BP; 382 A; 492 C; 421 G; 469 T; 0 other;

Query Match 51.5%; Score 282.4; DB 15; Length 1764;
Best Local Similarity 99.6%; Pred. No. 1.3e-66;
Matches 283; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    265 ATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCAT 324
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    461 ACGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCAT 520

Qy    325 TTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCA 384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 TTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCA 580
```

```
Qy     385 ATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGAT 444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 ATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGAT 640

Qy     445 ATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAG 504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 ATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAG 700

Qy     505 TCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
           |||||||||||||||||||||||||||||||||||||||||||
Db     701 TCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 744
```

RESULT 12
AAS68315
ID AAS68315 standard; cDNA; 772 BP.
XX
AC AAS68315;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4119.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG04128.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 4119; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 772 BP; 194 A; 191 C; 223 G; 164 T; 0 other;

Query Match 45.4%; Score 249; DB 23; Length 772;
Best Local Similarity 100.0%; Pred. No. 1.1e-57;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 155

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     156 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 215

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 275

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     276 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 335

Qy     241 GCAGGCCAG 249
           |||||||||
Db     336 GCAGGCCAG 344
```

RESULT 13
ABK70887
ID ABK70887 standard; cDNA; 1209 BP.
XX
AC ABK70887;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human cDNA encoding PAR1 type thrombin receptor delta 1-49TR.
XX
KW Human; ss; gene; PAR1; thrombin receptor; antiinflammatory; cytostatic;
KW inflammatory disease; cell proliferative disease.
XX
OS Homo sapiens.
XX
PN JP2002010784-A.
XX
PD 15-JAN-2002.
XX
PF 29-JUN-2000; 2000JP-0196514.
XX
PR 29-JUN-2000; 2000JP-0196514.
XX
PA (TEIJ) TEIJIN LTD.
XX
DR WPI; 2002-321520/36.
DR P-PSDB; ABG35298.
XX
PT An inhibitor of cell growth mediated by thrombin used to treat
PT inflammatory and cell proliferative diseases -
XX
PS Example 1; Page 22-24; 44pp; Japanese.
XX
CC The invention relates to a polypeptide or a compound which can inhibit
CC cell growth caused by thrombin. The polypeptide/compound combines to a
CC specific region of the structure of PAR1 type human thrombin receptor
CC participating to cell growth. Preferably, the compound contains the
CC 52nd to the 56th amino acid sequences at the amino end side of PAR1 type
CC human thrombin receptor ((X4)-Tyr-Glu-Pro-Phe-Trp-(X5) X4, X5 = optional
CC amino acid or peptide sequence). Also included are a modified PAR1 type
CC thrombin receptor gene or its fragment used for obtaining the above
CC polypeptide, a human PAR1 type thrombin receptor protein and its
CC encoding DNA comprising a fully. The polypeptide or the compound is used
CC to treat inflammatory diseases and cell proliferative diseases. The
CC present sequence is a cDNA encoding a PAR1 type thrombin receptor (or a
CC modified version).

```
XX
SQ   Sequence 1209 BP; 254 A; 330 C; 273 G; 352 T; 0 other;

  Query Match             43.9%;  Score 240.8;  DB 24;  Length 1209;
  Best Local Similarity   93.7%;  Pred. No. 2.3e-55;
  Matches  251;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy     281 CCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGA 340
           || |   ||   |  | | | ||     ||||||||||||||||||||||||||||||
Db      51 CCCGCTGTTGTCTGCCCGCACCCGGGCCGATAAATATGAACCATTTTGGGAGGATGAGGA 110

Qy     341 GAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCT 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 GAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCT 170

Qy     401 TCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTG 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171 TCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTG 230

Qy     461 GCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAA 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     231 GCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAA 290

Qy     521 CATCATGGCCATCGTTGTGTTCATCCTG 548
           ||||||||||||||||||||||||||||
Db     291 CATCATGGCCATCGTTGTGTTCATCCTG 318


RESULT 14
AAT93367
ID   AAT93367 standard; cDNA; 1312 BP.
XX
AC   AAT93367;
XX
DT   01-MAY-1998  (first entry)
XX
DE   Mouse thrombin receptor cDNA.
XX
KW   Transgenic animal; thrombin receptor; animal model; restenosis;
KW   knockout animal; antagonist; agonist; mouse; murine;
KW   Alzheimer's disease; wound repair; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   primer_bind     complement (612..631)
FT                   /*tag=  a
FT                   /label= MTR-3
FT   primer_bind     1294..1311
FT                   /*tag=  b
FT                   /label= MTR-2
XX
PN   WO9740138-A1.
XX
PD   30-OCT-1997.
XX
PF   22-APR-1997;   97WO-US06575.
XX
PR   25-JUN-1996;   96US-0020544.
PR   23-APR-1996;   96US-0016070.
XX
PA   (ORTH ) ORTHO PHARM CORP.
XX
PI   Andrade-Gordon P,  Leung W,  Zhou L;
XX
DR   WPI; 1997-535823/49.
XX
PT   Transgenic animals with altered form of thrombin receptor gene -
PT   useful in study of receptor function and identification of
PT   antagonists and agonists for disease therapy
XX
PS   Example 1; Fig 1; 47pp; English.
XX
CC   The present sequence was used in the preparation of a transgenic
CC   non-human vertebrate with somatic and germ cells containing an
CC   altered form of a thrombin receptor (TR) gene, where the altered
CC   gene has replaced a wild type TR gene in the animal or an ancestor
CC   at an embryonic stage using embryonic stem cells. The transgenic
CC   animal provides an animal model to understand receptor function in,
CC   e.g. platelets, endothelial cells and neurons, and evaluate drug
CC   therapies modulating TR function/expression in human cells.
CC   Alterations rendering the wild type gene non-functional produce
CC   knockout animals, useful as non-human whole animal models for
CC   diseases involving TR equivalents in humans. Alternatively,
CC   substituting the naturally occurring gene for a gene from a 2nd
CC   species or with a mutation, results in animals producing the 2nd
CC   species or mutated gene products. These transgenic animals are
CC   useful for drug antagonist and agonist studies, creation of animal
CC   models of human diseases and investigation of disorders associated
CC   with human TR-mediated responses. Antagonists may be used
CC   therapeutically for cardiovascular problems, e.g. restenosis, to
CC   limit platelet adhesion during interventional procedures or to
CC   generate therapies for bone degenerative and neurodegenerative
CC   diseases, e.g. Alzheimer's. Agonists may be useful in wound repair.
XX
SQ   Sequence 1312 BP; 247 A; 406 C; 328 G; 331 T; 0 other;

  Query Match             33.6%;  Score 184.2;  DB 18;  Length 1312;
  Best Local Similarity   67.8%;  Pred. No. 5.4e-40;
  Matches  280;  Conservative    0;  Mismatches  118;  Indels   15;  Gaps    1;

Qy     151 AGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCG 210
           ||||  |||||||||||||| |||||  |||||  | || |||  | || |||| |||||||
Db       4 AGCCTTGGACAATGGGGCCCCGGCGCTTGCTGATCGTCGCCCTCGGCCTCAGCCTGTGCG 63

Qy     211 GCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCA 270
           | ||  || ||||| |||||  ||       ||  ||||||||||  |    ||| |||| |
Db      64 GTCCCTTGCTGTCTTCCCGCGTCCCTATGAGCCAGCCAGAATCAGAGAGGACAGATGCTA 123

Qy     271 CCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATT----- 325
           |  |  | ||||| |||||  |||| |||||  |||| ||| ||          | |
Db     124 CGGTGAACCCCCGCTCATTCTTTCTAAGGAATCCCAGTGAAAATACATTTGAACTGGTCC 183

Qy     326 ----------TTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAG 375
                     |  |||||| |||||||||||||||||  |   |    ||    ||   ||
Db     184 CCCTGGGGGATGAGGAGGAGGAGGAGAAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAG 243

Qy     376 TCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATG 435
           ||| | | |||| ||||   |||| ||| |  |  | |||| | |||||||| || || |
Db     244 TCTACTTAAATATAAGCCTCCCTCCTCACACGCCGCCTCCTCCCTTCATCTCCGAGGACG 303

Qy     436 CCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAG 495
           |||||||||| |||||||| |||||||||| |||||  | || || ||||||||    |
Db     304 CCTCCGGATATCTGACCAGCCCCTGGCTGACGCTCTTCATGCCCTCCGTGTACACGATTG 363

Qy     496 TGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
           ||||  | |||||||| || || ||| || |||||||||  |||||| || ||
Db     364 TGTTCATTGTCAGCCTTCCTCTGAACGTCCTGGCCATCGCAGTGTTCGTCTTG 416


RESULT 15
ABL32681/c
ID   ABL32681 standard; DNA; 6436 BP.
XX
AC   ABL32681;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 654.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
```

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 654; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 6436 BP; 1509 A; 261 C; 1477 G; 3189 T; 0 other;

Query Match 27.0%; Score 148.2; DB 24; Length 6436;
Best Local Similarity 72.5%; Pred. No. 5.9e-30;
Matches 192; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
          |||||| ||||||  ||||| ||||||||||   |||||| |||||| ||| |||||
Db    1249 CGCCGAAAATCGCTTAAACCCTAATCTTACCCGTAAACACCCTACGCTCTACCTACCGCG 1190

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
          || |||| |||||||||||||| || |||  | | |   | || || | || |||||  |
Db    1189 AAAACCGACTCCCCGACCCGCAAAAATCAAAAAAAAAAATAAAACGAAACAACCCGAAAC 1130

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
          |   || ||||||| | || ||||||||| | |||| |||||    ||||| || || |
Db    1129 GAAACAACCTCCCGAAACAACGCCGCGCAAAACCCGAAACAATAAAACCGCGACGACTAC 1070

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
          | ||  |  |||||| ||||| ||| | || |||||| || ||| |||||||||||  |||
Db    1069 TACTAATAACCGCCTACTTCAATCTATACGACCCGCTATTATCTACCCGCACCCGAACCC 1010

Qy     241 GCAGGCCAGAATCAAAAGCAACAAA 265
          |||  ||| |   ||||   || ||
Db    1009 GCAAACCAAATAAAAAATACACGAA 985
```

Search completed: December 9, 2002, 04:16:37
Job time : 318 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 04:10:21 ; Search time 93 Seconds
(without alignments)
2297.668 Million cell updates/sec

Title: US-09-744-679-7
Perfect score: 548
Sequence: 1 cgccgagggtcgcttggacc.....ccatcgttgtgttcatcctg 548

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		44.6	8.1	192	9	US-09-925-664-33	Sequence 33, Appl
2		40.6	7.4	189	9	US-09-925-664-31	Sequence 31, Appl
3		40.2	7.3	32195	10	US-09-764-870-611	Sequence 611, App
4		40.2	7.3	32195	10	US-09-764-870-617	Sequence 617, App
5		40.2	7.3	32195	10	US-09-764-869-1605	Sequence 1605, Ap
6		39.6	7.2	409	10	US-09-960-352-13321	Sequence 13321, A
7	c	37.8	6.9	286	9	US-10-096-549-4	Sequence 4, Appli
8		37.6	6.9	2850	12	US-10-074-527-1	Sequence 1, Appli
9	c	37.6	6.9	14486	9	US-09-764-868-1504	Sequence 1504, Ap
10		37.4	6.8	1631	10	US-09-925-300-528	Sequence 528, App
11		37.4	6.8	43058	10	US-09-954-456-292	Sequence 292, App
12		37.4	6.8	43058	10	US-09-954-456-529	Sequence 529, App
13		37.4	6.8	43058	10	US-09-880-107-3950	Sequence 3950, Ap
14		37.2	6.8	193	10	US-09-864-761-21978	Sequence 21978, A
15		37.2	6.8	12425	12	US-10-023-529-50	Sequence 50, Appl
16		37.2	6.8	12425	12	US-10-023-523-50	Sequence 50, Appl
17		37	6.8	80959	9	US-09-858-546-3	Sequence 3, Appli
18	c	36.4	6.6	1407	10	US-09-833-381-1620	Sequence 1620, Ap
19		36	6.6	3715	10	US-09-880-107-2300	Sequence 2300, Ap
20		35.8	6.5	333	10	US-09-960-352-5606	Sequence 5606, Ap
21	c	35.8	6.5	2561	12	US-10-023-529-48	Sequence 48, Appl
22	c	35.8	6.5	2561	12	US-10-023-523-48	Sequence 48, Appl
23		35.6	6.5	1746	12	US-10-074-527-3	Sequence 3, Appli
24		35.4	6.5	930	9	US-09-960-631A-4	Sequence 4, Appli
25	c	35.2	6.4	4973	12	US-10-044-090-293	Sequence 293, App
26		35.2	6.4	5119	9	US-09-887-540A-1	Sequence 1, Appli
27	c	35	6.4	2538	9	US-10-136-573A-1	Sequence 1, Appli
28	c	35	6.4	2538	9	US-09-877-665-1	Sequence 1, Appli
29	c	35	6.4	2538	10	US-09-817-647-1	Sequence 1, Appli
30	c	34.8	6.4	1224	10	US-09-815-242-7867	Sequence 7867, Ap
31	c	34.8	6.4	2175	10	US-09-834-975-795	Sequence 795, App
32	c	34.8	6.4	88421	9	US-09-976-059-1	Sequence 1, Appli
33		34.6	6.3	405	10	US-09-864-761-6167	Sequence 6167, Ap
34		34.6	6.3	1110	10	US-09-893-737-49	Sequence 49, Appl
35	c	34.6	6.3	2750	10	US-09-905-983-51	Sequence 51, Appl
36	c	34.6	6.3	4000	10	US-09-954-456-2153	Sequence 2153, Ap
37	c	34.6	6.3	4000	10	US-09-880-107-3803	Sequence 3803, Ap
38	c	34.6	6.3	4826	10	US-09-772-304A-1	Sequence 1, Appli
39		34.6	6.3	42999	9	US-09-799-462A-17	Sequence 17, Appl
40		34.6	6.3	42999	9	US-10-125-767-17	Sequence 17, Appl
41		34.4	6.3	916	10	US-09-764-870-123	Sequence 123, App
42		34.4	6.3	1668	10	US-09-993-811-13	Sequence 13, Appl
43		34.4	6.3	2499	10	US-09-350-259-96	Sequence 96, Appl
44		34.4	6.3	3124	9	US-09-974-298-128	Sequence 128, App
45		34.4	6.3	3230	12	US-10-002-600-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1
US-09-925-664-33
; Sequence 33, Application US/09925664
; Patent No. US20020160006A1
; GENERAL INFORMATION:
; APPLICANT: Denney, Jr., Dan W.
; TITLE OF INVENTION: Methods of Treating Lymphoma and Leukemia
; FILE REFERENCE: GENITOPE-06499
; CURRENT APPLICATION NUMBER: US/09/925,664
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/370,453
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: 08/644,664
; PRIOR FILING DATE: 1996-05-01
; PRIOR APPLICATION NUMBER: 08/761,277
; PRIOR FILING DATE: 1996-12-06
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 192
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: CDS
; LOCATION: (1)..(189)
; OTHER INFORMATION:
US-09-925-664-33

Query Match 8.1%; Score 44.6; DB 9; Length 192;
Best Local Similarity 71.1%; Pred. No. 0.00093;
Matches 59; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy    273 TTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAG 332
          || ||||| || || |||||  |  | || || |||||||| || ||||||||||||||
Db      1 TTGGATCCACGATCGTTTCTATTGCGCAATCCAAATGATAAGTACGAACCATTTTGGGAA 60

Qy    333 GATGAGGAGAAAAATGAAAGTGG 355
          ||| ||   || |     |||||
Db     61 GATCAGAGCAAGATGCTGAGTGG 83
```

```
RESULT 2
US-09-925-664-31
; Sequence 31, Application US/09925664
; Patent No. US20020160006A1
; GENERAL INFORMATION:
;  APPLICANT: Denney, Jr., Dan W.
;  TITLE OF INVENTION: Methods of Treating Lymphoma and Leukemia
;  FILE REFERENCE: GENITOPE-06499
;  CURRENT APPLICATION NUMBER: US/09/925,664
;  CURRENT FILING DATE:  2001-08-09
;  PRIOR APPLICATION NUMBER: 09/370,453
;  PRIOR FILING DATE: 1999-08-09
;  PRIOR APPLICATION NUMBER: 08/644,664
;  PRIOR FILING DATE: 1996-05-01
;  PRIOR APPLICATION NUMBER: 08/761,277
;  PRIOR FILING DATE: 1996-12-06
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
;   LENGTH: 189
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
;   NAME/KEY: CDS
;   LOCATION: (1)..(186)
;   OTHER INFORMATION:
US-09-925-664-31

  Query Match             7.4%;  Score 40.6;  DB 9;  Length 189;
  Best Local Similarity   77.8%;  Pred. No. 0.013;
  Matches   49;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy     273 TTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAG 332
           || ||||| || || |||||  |  | || || |||||||| || ||||||||||||||
Db       1 TTGGATCCACGATCGTTTCTATTGCGCAATCCAAATGATAAGTACGAACCATTTTGGGAA 60

Qy     333 GAT 335
           |||
Db      61 GAT 63


RESULT 3
US-09-764-870-611
; Sequence 611, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14
;  CURRENT APPLICATION NUMBER: US/09/764,870
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
;   LENGTH: 32195
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-870-611

  Query Match             7.3%;  Score 40.2;  DB 10;  Length 32195;
  Best Local Similarity   48.9%;  Pred. No. 0.27;
  Matches  108;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy      30 CCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCTCCCCGACCCGCAGAAGTCA 89
           ||  ||   ||||  | |  | |  || |      | |||| | |  ||||||||
Db   22610 CCTCTGCCAACCCGCCCCGTTCCACGCAGGTGCTCCGGGCTGCACCGCCCGCAGACCGAG 22669

Qy      90 GGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCA 149
           |  |||  | || ||| ||||   | || || | ||  || || || ||   |  |
Db   22670 GCCGAGGTGCTGGAGCAGAGCGCGCAGACGCTGCGCGCCCACCTGGGGGCCCTGCTGAGC 22729

Qy     150 GAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGC 209
           | || | |      ||  |    | | |      |||||  ||||  |||| | | |  |
Db   22730 GCGCTCAGCCGCTCGGTTCGCGCGTGCCCCGCCGTGGTGCGCGCCACCTTCCGCCAGCTC 22789

Qy     210 GGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGA 250
             || ||   ||   |  |||  ||  | |  |  ||  ||
Db   22790 TTCCGGCGCGTGCGCGAGCGCTTCCCCGGCGCCCAGCACGA 22830


RESULT 4
US-09-764-870-617
; Sequence 617, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14
;  CURRENT APPLICATION NUMBER: US/09/764,870
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 617
;   LENGTH: 32195
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-870-617

  Query Match             7.3%;  Score 40.2;  DB 10;  Length 32195;
  Best Local Similarity   48.9%;  Pred. No. 0.27;
  Matches  108;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy      30 CCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCTCCCCGACCCGCAGAAGTCA 89
           ||  ||   ||||  | |  | |  || |      | |||| | |  ||||||||
Db   22610 CCTCTGCCAACCCGCCCCGTTCCACGCAGGTGCTCCGGGCTGCACCGCCCGCAGACCGAG 22669

Qy      90 GGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCA 149
           |  |||  | || ||| ||||   | || || | ||  || || || |   |  |
Db   22670 GCCGAGGTGCTGGAGCAGAGCGCGCAGACGCTGCGCGCCCACCTGGGGGCCCTGCTGAGC 22729

Qy     150 GAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGC 209
           | || | |      ||  |    | | |      |||||  ||||  |||| | | |  |
Db   22730 GCGCTCAGCCGCTCGGTTCGCGCGTGCCCCGCCGTGGTGCGCGCCACCTTCCGCCAGCTC 22789

Qy     210 GGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGA 250
             || ||   ||   |  |||  ||  | |  |  ||  ||
Db   22790 TTCCGGCGCGTGCGCGAGCGCTTCCCCGGCGCCCAGCACGA 22830


RESULT 5
US-09-764-869-1605
; Sequence 1605, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007
;  CURRENT APPLICATION NUMBER: US/09/764,869
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2442
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1605
;   LENGTH: 32195
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-869-1605

  Query Match             7.3%;  Score 40.2;  DB 10;  Length 32195;
  Best Local Similarity   48.9%;  Pred. No. 0.27;
  Matches  108;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy      30 CCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCTCCCCGACCCGCAGAAGTCA 89
```

```
Db  22610 CCTCTGCCAACCCGCCCCGTTCCACGCAGGTGCTCCGGGCTGCACCGCCCGCAGACCGAG 22669

Qy     90 GGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCA 149

Db  22670 GCCGAGGTGCTGGAGCAGAGCGCGCAGACGCTGCGCGCCCACCTGGGGGCCCTGCTGAGC 22729

Qy    150 GAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGC 209

Db  22730 GCGCTCAGCCGCTCGGTTCGCGCGTGCCCCGCCGTGGTGCGCGCCACCTTCCGCCAGCTC 22789

Qy    210 GGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGA 250

Db  22790 TTCCGGCGCGTGCGCGAGCGCTTCCCCGGCGCCCAGCACGA 22830


RESULT 6
US-09-960-352-13321
; Sequence 13321, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13321
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (297)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: 57-LIB3057-006-Q1-K1-G2
US-09-960-352-13321

  Query Match             7.2%;  Score 39.6;  DB 10;  Length 409;
  Best Local Similarity   49.7%;  Pred. No. 0.039;
  Matches   99;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy      29 ACCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCTCCCCGACCCGCAGAAGTC 88

Db     205 ACCCCCCCGTCCCCGGGCGCCGAGCGGCGATGAACTCGACCTTCCTGAACCACAGCGGCC 264

Qy      89 AGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGC 148

Db     265 TGGAGACAGCGGGTGGCTTGGGCGGCGGCCGCNGCGGGGCTGCGCTGGGGAACCGCAGGC 324

Qy     149 AGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTG 208

Db     325 ACGGCCTGGGCACGTGGCTGGGCTGCTGGCCCTGGGGTGCGCCGCTGTCCGCCAACGACG 384

Qy     209 CGGCCCGCTGTTGTCTGCC 227

Db     385 GGGTCCCCCGCGGGCTGGC 403


RESULT 7
US-10-096-549-4/c
; Sequence 4, Application US/10096549
; Patent No. US20020165191A1
;   GENERAL INFORMATION:
;        APPLICANT: Moonen, Chrit
;        TITLE OF INVENTION: Spatial and Temporal Control of Gene
;                            Expression Using a Heat Shock Protein Promoter in
;                            Combination with Local Heat
;        NUMBER OF SEQUENCES: 7
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/096,549
;             FILING DATE: 11-Mar-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/242,244
;             FILING DATE: 29-Feb-2000
;             APPLICATION NUMBER: WO PCT/US97/15270
;             FILING DATE: 14-AUG-1997
;             APPLICATION NUMBER: US 60/024,213
;             FILING DATE: 15-AUG-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Weber, Kenneth A.
;             REGISTRATION NUMBER: 31,677
;             REFERENCE/DOCKET NUMBER: 15280-2641PC
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 576-0200
;             TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 286 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  -
;             LOCATION:  1..286
;             OTHER INFORMATION:  /note= "human heat shock protein
;             (hsp70B) promoter"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-096-549-4

  Query Match             6.9%;  Score 37.8;  DB 9;  Length 286;
  Best Local Similarity   60.0%;  Pred. No. 0.11;
  Matches   63;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      69 CTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGC 128

Db     227 CTTCCCGCACCTTCCCGCCCAGTCGAGAGCCTCCCGCCCGCCCGCCCGGGCTGAGTCAGC 168

Qy     129 CTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGG 173

Db     167 CGGGCGGCGCAGCGCGGCGCTGACCCCGGGAGAAGGTTCGGGCGG 123


RESULT 8
US-10-074-527-1
; Sequence 1, Application US/10074527
; Patent No. US20020142426A1
; GENERAL INFORMATION:
;  APPLICANT: Olandt, Peter J.
;  APPLICANT:  Meyers, Rachel E.
;  APPLICANT:  Galvin, Katherine A.
;  APPLICANT:  Millennium Pharmaceuticals Inc.
;  TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
;  TITLE OF INVENTION:  Uses Therefor
;  FILE REFERENCE: MPI2001-018P1RCP1(M)
;  CURRENT APPLICATION NUMBER: US/10/074,527
;  CURRENT FILING DATE:  2002-02-12
;  PRIOR APPLICATION NUMBER: 60/269202
```

```
;  PRIOR FILING DATE: 2001-02-15
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2850
;   TYPE: DNA
;   ORGANISM: homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (81)...(1826)
US-10-074-527-1

  Query Match             6.9%;  Score 37.6;  DB 12;  Length 2850;
  Best Local Similarity   51.8%;  Pred. No. 0.42;
  Matches   85;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy      90 GGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCA 149
           || | | |  ||  | || |||   || ||||| || |||  | ||| | ||| ||||
Db      72 GGCGGGCGCATGTGGGGGCGCACGGCGCGGCGGCGCTGCCCGCGGGAACTGCGGCGCGGC 131

Qy     150 GAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGC 209
             |   | |    ||| ||  | |  ||| |||  | ||||||       |    ||
Db     132 CGGGAGGCGCTGTTGGTGCTCCTGGCGCTACTGGCGTTGGCCGGGCTGGGCTCGGTGCTG 191

Qy     210 GGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATC 253
            |  ||| |      ||| |  || ||||   |  ||| | | |
Db     192 CGGGCGCAGCGTGGGGCCGGGGCCGGGGCTGCCGAGCCGGGACC 235


RESULT 9
US-09-764-868-1504/c
; Sequence 1504, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ32
;  CURRENT APPLICATION NUMBER: US/09/764,868
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1510
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1504
;   LENGTH: 14486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-868-1504

  Query Match             6.9%;  Score 37.6;  DB 9;  Length 14486;
  Best Local Similarity   53.8%;  Pred. No. 1;
  Matches   99;  Conservative    0;  Mismatches   84;  Indels    1;  Gaps    1;

Qy      66 CGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGC 125
           |||  |||||| | |||| || ||||  |||| | |  | |      |  ||| |  |
Db    4230 CGGGCCCCCGAACTGCAGGAGGCAGG-TAGAGCGGGGCGTGAGCGCCCGGGAGCCTCGCT 4172

Qy     126 AGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTG 185
             |||||||| |  | |  | |||  | ||   || ||| ||  ||||||||||||  |
Db    4171 GCGCTCCCGGACCCTCCCTCCCCAGCCCGCGCCTCACTGGAGCTCCGGCGGCTGCTGAGG 4112

Qy     186 GTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGG 245
              | | ||||        ||  |      |  ||| |   | ||| ||| | |||
Db    4111 CGCAGCAGCGGCTTGGACACCAGCTCCGACTCCTCTTCTTCGTCCTCCCCGGCGCTCAGC 4052

Qy     246 CCAG 249
            | |
Db    4051 TCCG 4048


RESULT 10
US-09-925-300-528
; Sequence 528, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 528
;   LENGTH: 1631
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (1628)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-528

  Query Match             6.8%;  Score 37.4;  DB 10;  Length 1631;
  Best Local Similarity   60.2%;  Pred. No. 0.36;
  Matches   62;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy     118 GGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGC 177
           |||| || ||||  || |||| | |||| || |||||  |    |  |||| ||| |||
Db      39 GGCGCGGGAGCCGACCCGAGCCGAGCCGAGCCGAGCCGAGCCGGAGCGGGCGGCGAAGGC 98

Qy     178 TGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTT 220
             |  || |   || ||   | |     ||| |||   ||||||
Db      99 CGGCGCGGCGAGCAGCAACCATGTCGGTGTTCGGGAAGCTGTT 141


RESULT 11
US-09-954-456-292
; Sequence 292, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 292
;   LENGTH: 43058
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

US-09-954-456-292

Query Match 6.8%; Score 37.4; DB 10; Length 43058;
Best Local Similarity 51.7%; Pred. No. 2;
Matches 109; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

```
Qy      11 CGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCT 70
           ||| ||||||| |       |   | || |   ||||  |||  || | |  |  |||
Db   25526 CGCGTGGACCCGGTCAACTTCAAGGTGCGCGGGGCGCGGTGCGGGCGGGGCGGGGCGGGG 25585

Qy      71 CCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCT 130
           || ||   ||     | | |  | | ||| |  |||| || |  || || ||||  ||
Db   25586 CCGCGGGGCGGGCGGGGCCGCGGGGCGGG-GTCGCGGGGCGGGGCGGGGTGGGGTCGCGG 25644

Qy     131 CCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGC 190
             ||| || | | ||||  | |    ||     || |  ||||  |  || || || | |
Db   25645 GGCGGGGCGGGGTCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGCGGGGCGGCCGGGGCC 25704

Qy     191 CGCCTGCTTCAGTCTGTGCGGCCCGCTGTTG 221
           || | |     | | | ||||   |  | ||
Db   25705 CGGCGGGGCGGGGCGGGGCGGGGAGGGGCTG 25735
```

RESULT 12
US-09-954-456-529
; Sequence 529, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 529
; LENGTH: 43058
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-529

Query Match 6.8%; Score 37.4; DB 10; Length 43058;
Best Local Similarity 51.7%; Pred. No. 2;
Matches 109; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

```
Qy      11 CGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCT 70
           ||| ||||||| |       |   | || |   ||||  |||  || | |  |  |||
Db   25526 CGCGTGGACCCGGTCAACTTCAAGGTGCGCGGGGCGCGGTGCGGGCGGGGCGGGGCGGGG 25585

Qy      71 CCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCT 130
           || ||   ||     | | |  | | ||| |  |||| || |  || || ||||  ||
Db   25586 CCGCGGGGCGGGCGGGGCCGCGGGGCGGG-GTCGCGGGGCGGGGCGGGGTGGGGTCGCGG 25644

Qy     131 CCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGC 190
             ||| || | | ||||  | |    ||     || |  ||||  |  || || || | |
Db   25645 GGCGGGGCGGGGTCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGCGGGGCGGCCGGGGCC 25704

Qy     191 CGCCTGCTTCAGTCTGTGCGGCCCGCTGTTG 221
           || | |     | | | ||||   |  | ||
Db   25705 CGGCGGGGCGGGGCGGGGCGGGGAGGGGCTG 25735
```

Qy 131 CCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGC 190
Db 25645 GGCGGGGCGGGGTCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGCGGGGCGGCCGGGGCC 25704

Qy 191 CGCCTGCTTCAGTCTGTGCGGCCCGCTGTTG 221
Db 25705 CGGCGGGGCGGGGCGGGGCGGGGAGGGGCTG 25735

RESULT 13
US-09-880-107-3950
; Sequence 3950, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3950
; LENGTH: 43058
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z84721
US-09-880-107-3950

Query Match 6.8%; Score 37.4; DB 10; Length 43058;
Best Local Similarity 51.7%; Pred. No. 2;
Matches 109; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

```
Qy      11 CGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCT 70
           ||| ||||||| |       |   | || |   ||||  |||  || | |  |  |||
Db   25526 CGCGTGGACCCGGTCAACTTCAAGGTGCGCGGGGCGCGGTGCGGGCGGGGCGGGGCGGGG 25585

Qy      71 CCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCT 130
           || ||   ||     | | |  | | ||| |  |||| || |  || || ||||  ||
Db   25586 CCGCGGGGCGGGCGGGGCCGCGGGGCGGG-GTCGCGGGGCGGGGCGGGGTGGGGTCGCGG 25644

Qy     131 CCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGC 190
             ||| || | | ||||  | |    ||     || |  ||||  |  || || || | |
Db   25645 GGCGGGGCGGGGTCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGCGGGGCGGCCGGGGCC 25704

Qy     191 CGCCTGCTTCAGTCTGTGCGGCCCGCTGTTG 221
           || | |     | | | ||||   |  | ||
Db   25705 CGGCGGGGCGGGGCGGGGCGGGGAGGGGCTG 25735
```

RESULT 14
US-09-864-761-21978
; Sequence 21978, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312

```
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21978
;  LENGTH: 193
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AC004084.1
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.8
;  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
;  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.7
;  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
;  OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
;  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.7
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.1
;  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.7
;  OTHER INFORMATION: NT HIT: gi11419196, EVALUE 1.00e-105
;  OTHER INFORMATION: EST_HUMAN HIT: BE207240.1, EVALUE 2.00e-78
;  OTHER INFORMATION: SWISSPROT HIT: P40823, EVALUE 2.90e+00
US-09-864-761-21978

  Query Match             6.8%;  Score 37.2;  DB 10;  Length 193;
  Best Local Similarity   50.0%;  Pred. No. 0.13;
  Matches   93;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      65 CCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGG 124
           | |||| | |  ||||||||     |  |||  | || ||| ||||   | || || | |
Db       7 CGGGCTGCACCGCCCGCAGACCGAGGCCGAGGTGCTGGAGCAGAGCGCGCAGACGCTGCG 66

Qy     125 CAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCT 184
           |  || || || |   |  |.    | || | |      ||  |    | | |      |
Db      67 CGCCCACCTGGGGGCCCTGCTGAGCGCGCTCAGCCGCTCGGTTCGCGCGTGCCCCGCCGT 126

Qy     185 GGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAG 244
           ||||  ||||  |||| | | |  |  || ||   ||   |  |||  ||  | | |
Db     127 GGTGCGCGCCACCTTCCGCCAGCTCTTCCGGCGCGTGCGCGAGCGCTTCCCCGGCGCCCA 186

Qy     245 GCCAGA 250
```

```
           || ||
Db    187 GCACGA 192


RESULT 15
US-10-023-529-50
; Sequence 50, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/023,529
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: 09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
;  LENGTH: 12425
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-023-529-50

  Query Match             6.8%;  Score 37.2;  DB 12;  Length 12425;
  Best Local Similarity   48.6%;  Pred. No. 1.2;
  Matches  102;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy      44 GCGCTCTGCCTGCCGCGAAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAA 103
           | || | | || |  |  | |   |  | |      |  | ||  |||| | ||| |
Db    2333 GGGCCCGGGCTCCACCTTAAAAGCGGGCGCGTGGGGGTGGGAGGGAGGAAGGCGGGCGGC 2392

Qy     104 GCGGAGCAGCCCGAGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAAT 163
           | |||| ||   | |  ||           |  | || | | ||        | |    |
Db    2393 GGGGAGGAGGGAGGGAGGGAAGGAAGGGGGGCCGGAGTGTCCCGGGCGCAGGGCGCGCGT 2452

Qy     164 GGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTC 223
           | ||| ||||||||| || |   | ||||| | ||  |   ||   |  | | |  | |
Db    2453 GCGGCGGCGGCGGCGGCGGGGAGGGGCCGGCCGCGCCGCGCTCCCCTCCTCCCCCTCGCA 2512

Qy     224 TGCCCGCACCCGGGCCCGCAGGCCAGAATC 253
           | ||||  |||| || |||    ||||| |
Db    2513 TCCCCGGCCCCGCGCGCGCCCAGCAGAAGC 2542


Search completed: December  9, 2002, 05:54:03
Job time : 248 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 04:06:36 ; Search time 2427 Seconds
(without alignments)
3656.830 Million cell updates/sec

Title: US-09-744-679-7
Perfect score: 548
Sequence: 1 cgccgagggtcgcttggacc.....ccatcgttgtgttcatcctg 548

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	548	100.0	888	14	BQ440574	BQ440574 AGENCOURT
2	536	97.8	699	10	BE258615	BE258615 601106874
3	505.2	92.2	739	12	BG473471	BG473471 602516274
4	197	35.9	496	11	BC016059	BC016059 Homo sapi
5	183	33.4	621	10	BB662299	BB662299 BB662299
6	183	33.4	674	13	BI685426	BI685426 603308867
7	183	33.4	942	14	BQ896389	BQ896389 AGENCOURT
8	182.4	33.3	949	14	BQ923464	BQ923464 AGENCOURT
9	180.8	33.0	584	13	BI738768	BI738768 603362614
10	175.8	32.1	688	9	AA882290	AA882290 vx43b01.r
11	165.6	30.2	672	12	BE911999	BE911999 601662765
12	162.2	29.6	763	13	BI666195	BI666195 603287191
13	159.2	29.1	467	10	BB862683	BB862683 BB862683
14	157.6	28.8	466	10	BB852109	BB852109 BB852109
15	152	27.7	1429	14	BQ926183	BQ926183 AGENCOURT
16	145.8	26.6	405	12	BF116952	BF116952 uy92h02.y
17	133	24.3	557	12	BG067569	BG067569 H3055G12-
18	132	24.1	554	14	C87335	C87335 C87335 Mous
19	132	24.1	588	14	C86764	C86764 C86764 Mous
20	128	23.4	504	9	AA690768	AA690768 vt25e01.r
21	126.2	23.0	486	10	BE128304	BE128304 DEPA2054
c 22	126	23.0	350	12	BF407166	BF407166 UI-R-BJ2-
23	124.6	22.7	475	9	AA096771	AA096771 mo06h02.r
24	122.2	22.3	462	12	BG744398	BG744398 602723136
25	118.2	21.6	906	13	BI729686	BI729686 603349374
26	115	21.0	436	10	BB864321	BB864321 BB864321
27	94.8	17.3	370	10	BB866629	BB866629 BB866629
28	90.6	16.5	316	10	BE126861	BE126861 DEPA0609
29	85.8	15.7	273	12	BE864515	BE864515 UI-M-BH1-
30	84	15.3	375	10	BE127249	BE127249 DEPA0998
31	72.4	13.2	272	10	BB838066	BB838066 BB838066
32	70.2	12.8	326	10	BB865249	BB865249 BB865249
33	67.8	12.4	683	9	AJ449531	AJ449531 AJ449531
34	67.8	12.4	704	9	AJ449546	AJ449546 AJ449546
35	67.8	12.4	824	9	AJ456565	AJ456565 AJ456565
36	67.4	12.3	718	9	AJ447241	AJ447241 AJ447241
37	66.2	12.1	798	9	AJ455565	AJ455565 AJ455565
38	57.2	10.4	127	14	W20934	W20934 mb96c06.r1
c 39	51.6	9.4	925	17	CNS0091P	AL053013 Drosophil
c 40	51	9.3	910	17	CNS006ON	AL065629 Drosophil
41	51	9.3	925	17	CNS0091P	AL053013 Drosophil
42	50.4	9.2	935	17	CNS006XK	AL066051 Drosophil
43	49.6	9.1	1212	14	BQ069498	BQ069498 AGENCOURT
c 44	48.4	8.8	932	17	CNS0072Q	AL066742 Drosophil
c 45	48.2	8.8	1010	17	CNS016TE	AL107180 Drosophil

ALIGNMENTS

RESULT 1
BQ440574
LOCUS BQ440574 888 bp mRNA linear EST 24-MAY-2002
DEFINITION AGENCOURT_7907353 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6154861 5', mRNA sequence.
ACCESSION BQ440574
VERSION BQ440574.1 GI:21179650
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 888)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13496 row: j column: 14
High quality sequence stop: 654.
FEATURES Location/Qualifiers
source 1. .888

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6154861"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."

BASE COUNT 181 a 263 c 237 g 207 t
ORIGIN

Query Match 100.0%; Score 548; DB 14; Length 888;
Best Local Similarity 100.0%; Pred. No. 1.1e-127;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 84

Qy     61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 144

Qy    121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 204

Qy    181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 264

Qy    241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 324

Qy    301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 384

Qy    361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 444

Qy    421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 504

Qy    481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 564

Qy    541 TCATCCTG 548
          ||||||||
Db    565 TCATCCTG 572
```

RESULT 2
BE258615
LOCUS BE258615 699 bp mRNA linear EST 13-JUL-2000
DEFINITION 601106874F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3343051 5', mRNA sequence.
ACCESSION BE258615
VERSION BE258615.1 GI:9129107
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 699)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM127 row: c column: 20.
FEATURES Location/Qualifiers
source 1. .699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3343051"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 153 a 202 c 194 g 149 t 1 others
ORIGIN

Query Match 97.8%; Score 536; DB 10; Length 699;
Best Local Similarity 99.6%; Pred. No. 1.1e-124;
Matches 547; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy      1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 110

Qy     61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCC-GAGG 119
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    111 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCTGAGG 170

Qy    120 CGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTG 179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 CGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTG 230

Qy    180 CTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCC 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 CTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCC 290

Qy    240 CGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGG 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 CGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGG 350

Qy    300 AACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTA 359
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    351 AACCCCAATGATAAATATGAACCATTNTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTA 410

Qy    360 ACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCA 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 ACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCA 470

Qy    420 TTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCA 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 TTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCA 530

Qy    480 TCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTG 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 TCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTG 590

Qy    540 TTCATCCTG 548
          |||||||||
Db    591 TTCATCCTG 599
```

```
RESULT 3
BG473471
LOCUS       BG473471                739 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602516274F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4647883 5',
            mRNA sequence.
ACCESSION   BG473471
VERSION     BG473471.1  GI:13405746
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 739)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1425  row: c  column: 20
            High quality sequence stop: 718.
FEATURES             Location/Qualifiers
     source          1. .739
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4647883"
                     /clone_lib="NIH_MGC_16"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      162 a    207 c    200 g    170 t
ORIGIN

  Query Match             92.2%;  Score 505.2;  DB 12;  Length 739;
  Best Local Similarity   99.1%;  Pred. No. 6.6e-117;
  Matches  529;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy      16 GGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCTCCCCG 75
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCGAAGACCGGCTCCCCG 61

Qy      76 ACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCTCCCGG 135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 ACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCTCCCGG 121

Qy     136 AGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCT 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 AGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCT 181

Qy     196 GCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAA 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAA 241

Qy     256 AAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAAT 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 AAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAAT 301

Qy     316 ATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAG 375
           |||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     302 ATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAAGTGAATACAGATTAG 361

Qy     376 TCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCC-TGCATTCATCTCAGAAGAT 434
           ||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db     362 TCTCCATCAGTAAAAGCAGTCCTCTTCAAAAACAACTTCCTTGCATTCATCTCAGAAGAT 421

Qy     435 GCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGA 494
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     422 GCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTGGTCCCATCTGTGTACACCGGA 481

Qy     495 GTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 GTG-TTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 534


RESULT 4
BC016059
LOCUS       BC016059                496 bp    mRNA    linear   HTC 24-OCT-2001
DEFINITION  Homo sapiens, Similar to coagulation factor II (thrombin) receptor,
            clone IMAGE:4849569, mRNA.
ACCESSION   BC016059
VERSION     BC016059.1  GI:16359189
KEYWORDS    HTC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 34 Row: a Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6031164
            This clone has the following problem: incomplete processing.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4849569"
                     /tissue_type="Blood, natural killer cell"
                     /clone_lib="NIH_MGC_106"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
BASE COUNT      130 a    144 c    146 g     76 t
ORIGIN

  Query Match             35.9%;  Score 197;  DB 11;  Length 496;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-39;
  Matches  197;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      53 CTGCCGCGAAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAG 112
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGCCGCGAAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAG 60

Qy     113 CCCGAGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCG 172
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCCGAGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCG 120

Qy     173 GCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCAC 232
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCAC 180

Qy     233 CCGGGCCCGCAGGCCAG 249
           |||||||||||||||||
Db     181 CCGGGCCCGCAGGCCAG 197
```

RESULT 5
BB662299
LOCUS BB662299 621 bp mRNA linear EST 26-OCT-2001
DEFINITION BB662299 RIKEN full-length enriched, 16 days neonate heart Mus musculus cDNA clone D830039D11 5', mRNA sequence.
ACCESSION BB662299
VERSION BB662299.1 GI:16496054
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 621)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .621
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D830039D11"
/clone_lib="RIKEN full-length enriched, 16 days neonate heart"
/tissue_type="heart"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 116 a 197 c 173 g 135 t
ORIGIN

Query Match 33.4%; Score 183; DB 10; Length 621;
Best Local Similarity 65.3%; Pred. No. 1.2e-35;
Matches 292; Conservative 0; Mismatches 140; Indels 15; Gaps 1;

```
Qy     117 AGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGG 176
           ||  || |  | |   ||  |   |||  |     |  |||||||||||||| |||||
Db      52 AGTAGGCGACGGCGGGCGCCGGGCCGCGCGGGCAGCCTTGGGACAATGGGGCCCCGGCGC 111

Qy     177 CTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGG 236
           ||||| | || |||  | || |||| |||||||| ||  || ||||| |||||  ||
Db     112 TTGCTGATCGTCGCCCTCGGCCTCAGCCTGTGCGGTCCCTTGCTGTCTTCCCGCGTCCCT 171

Qy     237 GCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTC 296
           ||  |||||||||| |    ||| |||| ||  |  | |||||  |||||  ||||
Db     172 ATGAGCCAGCCAGAATCAGAGAGGACAGATGCTACGGTGAACCCCCGCTCATTCTTTCTA 231

Qy     297 AGGAACCCCAATGATAAATATGAACCATT--------------TTGGGAGGATGAGGAG 341
           ||||| |||| ||| ||          | |              |  |||||| ||||||
Db     232 AGGAATCCCAGTGAAAATACATTTGAACTGGTCCCCCTGGGGGATGAGGAGGAGGAGGAG 291

Qy     342 AAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTT 401
           ||||||||||| |   |    ||    ||||| | | |||| ||||   |||| |
Db     292 AAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAGTCTACTTAAATATAAGCCTCCCTCCT 351

Qy     402 CAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGG 461
           || |  | | |||| | |||||||| || || ||||||||||| |||||||| |||||
Db     352 CACACGCCGCCTCCTCCCTTCATCTCCGAGGACGCCTCCGGATATCTGACCAGCCCCTGG 411

Qy     462 CTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAAC 521
           ||||| |||||  | || || ||||||||    |||||  | |||||||| || || |||
Db     412 CTGACGCTCTTCATGCCCTCCGTGTACACGATTGTGTTCATTGTCAGCCTTCCTCTGAAC 471

Qy     522 ATCATGGCCATCGTTGTGTTCATCCTG 548
           || |||||||||  ||||||  || ||
Db     472 GTCCTGGCCATCGCAGTGTTCGTCTTG 498
```

RESULT 6
BI685426
LOCUS BI685426 674 bp mRNA linear EST 18-SEP-2001
DEFINITION 603308867F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5344888 5', mRNA sequence.
ACCESSION BI685426
VERSION BI685426.1 GI:15648054
KEYWORDS EST.

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11876  row: e  column: 17
            High quality sequence stop: 672.
FEATURES             Location/Qualifiers
     source          1. .674
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5344888"
                     /clone_lib="NCI_CGAP_Mam6"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      126 a    212 c    188 g    148 t
ORIGIN
```

```
  Query Match             33.4%;  Score 183;  DB 13;  Length 674;
  Best Local Similarity   65.3%;  Pred. No. 1.2e-35;
  Matches  292;  Conservative    0;  Mismatches  140;  Indels   15;  Gaps    1;

Qy    117 AGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGG 176
          ||  || |  | |   ||  |  |||   |     | ||||||||||||| |||||
Db     37 AGTAGGCGACGGCGGGCGCCGGGCCGCGCGGGCAGCCTTGGGACAATGGGGCCCCGGCGC 96

Qy    177 CTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGG 236
           ||||| | || |||  | || |||| |||||||| ||  || |||||  |||||  ||
Db     97 TTGCTGATCGTCGCCCTCGGCCTCAGCCTGTGCGGTCCCTTGCTGTCTTCCCGCGTCCCT 156

Qy    237 GCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTC 296
             ||  |||||||||| |    ||| |||| ||  |  | |||||  |||||  ||||
Db    157 ATGAGCCAGCCAGAATCAGAGAGGACAGATGCTACGGTGAACCCCCGCTCATTCTTTCTA 216

Qy    297 AGGAACCCCAATGATAAATATGAACCATT---------------TTGGGAGGATGAGGAG 341
          ||||| |||| ||| ||          | |               |  ||||||  ||||||
Db    217 AGGAATCCCAGTGAAAATACATTTGAACTGGTCCCCCTGGGGGATGAGGAGGAGGAGGAG 276

Qy    342 AAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTT 401
          ||||||||||| |   |    ||    ||   ||||| | | |||| ||||   |||| |
Db    277 AAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAGTCTACTTAAATATAAGCCTCCCTCCT 336

Qy    402 CAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGG 461
          || |  |  | |||| | |||||||| || || |||||||||||| |||||||| |||||
Db    337 CACACGCCGCCTCCTCCCTTCATCTCCGAGGACGCCTCCGGATATCTGACCAGCCCCTGG 396

Qy    462 CTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAAC 521
          ||||| |||||  | || || ||||||||    |||||  | |||||||| || || |||
Db    397 CTGACGCTCTTCATGCCCTCCGTGTACACGATTGTGTTCATTGTCAGCCTTCCTCTGAAC 456

Qy    522 ATCATGGCCATCGTTGTGTTCATCCTG 548
           || |||||||||  |||||| || ||
Db    457 GTCCTGGCCATCGCAGTGTTCGTCTTG 483
```

```
RESULT 7
BQ896389
LOCUS       BQ896389                942 bp    mRNA    linear   EST 16-AUG-2002
DEFINITION  AGENCOURT_8743817 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6389748
            5', mRNA sequence.
ACCESSION   BQ896389
VERSION     BQ896389.1  GI:22288403
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 942)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13875  row: e  column: 13
            High quality sequence stop: 697.
FEATURES             Location/Qualifiers
     source          1. .942
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6389748"
                     /clone_lib="NIH_MGC_129"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: olfactory epithelium; Vector:
                     pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                     unidirectionally. Primer: Oligo dT. Average insert size
                     2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                     is a NIH_MGC Library."
BASE COUNT      170 a    294 c    259 g    219 t
ORIGIN
```

```
  Query Match             33.4%;  Score 183;  DB 14;  Length 942;
  Best Local Similarity   65.3%;  Pred. No. 1.3e-35;
  Matches  292;  Conservative    0;  Mismatches  140;  Indels   15;  Gaps    1;

Qy    117 AGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGG 176
          ||  || |  | |   ||  |  |||   |     | ||||||||||||| |||||
Db     24 AGTAGGCGACGGCGGGCGCCGGGCCGCGCGGGCAGCCTTGGGACAATGGGGCCCCGGCGC 83

Qy    177 CTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGG 236
           ||||| | || |||  | || |||| |||||||| ||  || |||||  |||||  ||
Db     84 TTGCTGATCGTCGCCCTCGGCCTCAGCCTGTGCGGTCCCTTGCTGTCTTCCCGCGTCCCT 143

Qy    237 GCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTC 296
             ||  |||||||||| |    ||| |||| ||  |  | |||||  |||||  ||||
Db    144 ATGAGCCAGCCAGAATCAGAGAGGACAGATGCTACGGTGAACCCCCGCTCATTCTTTCTA 203

Qy    297 AGGAACCCCAATGATAAATATGAACCATT---------------TTGGGAGGATGAGGAG 341
          |||||  |||| ||| ||          |  |               |  ||||||  ||||||
Db    204 AGGAATCCCAGTGAAAATACATTTGAACTGGTCCCCCTGGGGGATGAGGAGGAGGAGGAG 263

Qy    342 AAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTT 401
          ||||||||||| |   |    ||    ||   ||||| | | |||| ||||   |||| |
Db    264 AAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAGTCTACTTAAATATAAGCCTCCCTCCT 323

Qy    402 CAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGG 461
          || |  |  | |||| | |||||||| || || |||||||||||| |||||||| |||||
Db    324 CACACGCCGCCTCCTCCCTTCATCTCCGAGGACGCCTCCGGATATCTGACCAGCCCCTGG 383

Qy    462 CTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAAC 521
          |||||  |||||  | || || ||||||||    |||||  | |||||||| || || |||
```

```
Db    384 CTGACGCTCTTCATGCCCTCCGTGTACACGATTGTGTTCATTGTCAGCCTTCCTCTGAAC 443

Qy    522 ATCATGGCCATCGTTGTGTTCATCCTG 548
          || ||||||||| |||||| || ||
Db    444 GTCCTGGCCATCGCAGTGTTCGTCTTG 470


RESULT 8
BQ923464
LOCUS       BQ923464                949 bp    mRNA    linear   EST 20-AUG-2002
DEFINITION  AGENCOURT_8800954 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6392393
            5', mRNA sequence.
ACCESSION   BQ923464
VERSION     BQ923464.1  GI:22338495
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 949)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13882  row: c  column: 18
            High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1. .949
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6392393"
                     /clone_lib="NIH_MGC_130"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: otocysts; Vector: pCMV-SPORT6.1.ccdb;
                     Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
                     Primer: Oligo dT.  Average insert size 1.95  kb.
                     Constructed by ResGen, Invitrogen Corp. Note: this is a
                     NIH_MGC Library."
BASE COUNT      161 a    315 c    261 g    212 t
ORIGIN

  Query Match             33.3%;  Score 182.4;  DB 14;  Length 949;
  Best Local Similarity   67.9%;  Pred. No. 1.9e-35;
  Matches  277;  Conservative    0;  Mismatches  116;  Indels   15;  Gaps    1;

Qy    156 GGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCG 215
          ||||||||||||||| |||||  |||||  | || |||  | || |||| |||||||| ||
Db     41 GGGACAATGGGGCCCCGGCGCTTGCTGATCGTCGCCCTCGGCCTCAGCCTGTGCGGTCCC 100

Qy    216 CTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTA 275
          || ||||| |||||  ||      ||  |||||||||| |    ||| |||| ||  |
Db    101 TTGCTGTCTTCCCGCGTCCCTATGAGCCAGCCAGAATCAGAGAGGACAGATGCTACGGTG 160

Qy    276 GATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATT---------- 325
          | ||||| |||||  |||| |||||  |||| ||| ||         | |
Db    161 AACCCCCGCTCATTCTTTCTAAGGAATCCCAGTGAAAATACATTTGAACTGGTCCCCCTG 220

Qy    326 -----TTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCC 380
               |  |||||| ||||||||||||||||| |   |    ||    ||   ||||| |
Db    221 GGGGATGAGGAGGAGGAGGAGAAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAGTCTAC 280

Qy    381 ATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCC 440
          | ||||  ||||   |||| ||| |  |  | |||| | |||||||| || || ||||||
Db    281 TTAAATATAAGCCTCCCTCCTCACACGCCGCCTCCTCCCTTCATCTCCGAGGACGCCTCC 340

Qy    441 GGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTT 500
          ||||||  |||||||| |||||||||| |||||  | || || ||||||||    |||||
Db    341 GGATATCTGACCAGCCCCTGGCTGACGCTCTTCATGCCCTCCGTGTACACGATTGTGTTC 400

Qy    501 GTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
          | |||||||| || || ||| || |||||||||  |||||| || ||
Db    401 ATTGTCAGCCTTCCTCTGAACGTCCTGGCCATCGCAGTGTTCGTCTTG 448


RESULT 9
BI738768
LOCUS       BI738768                584 bp    mRNA    linear   EST 20-SEP-2001
DEFINITION  603362614F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5369618 5',
            mRNA sequence.
ACCESSION   BI738768
VERSION     BI738768.1  GI:15715781
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11940  row: l  column: 03
            High quality sequence stop: 582.
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5369618"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      105 a    181 c    168 g    130 t
ORIGIN

  Query Match             33.0%;  Score 180.8;  DB 13;  Length 584;
  Best Local Similarity   67.6%;  Pred. No. 4.2e-35;
  Matches  276;  Conservative    0;  Mismatches  117;  Indels   15;  Gaps    1;

Qy    156 GGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCG 215
          ||||||||||||||| |||||  ||| | | || |||  | || |||| |||||||| ||
Db     36 GGGACAATGGGGCCCCGGCGCTTGCGGATCGTCGCCCTCGGCCTCAGCCTGTGCGGTCCC 95

Qy    216 CTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTA 275
          || ||||| |||||  ||      ||  |||||||||| |    ||| |||| ||  |
Db     96 TTGCTGTCTTCCCGCGTCCCTATGAGCCAGCCAGAATCAGAGAGGACAGATGCTACGGTG 155

Qy    276 GATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCATT---------- 325
          | ||||| |||||  |||| |||||  |||| ||| ||         | |
Db    156 AACCCCCGCTCATTCTTTCTAAGGAATCCCAGTGAAAATACATTTGAACTGGTCCCCCTG 215

Qy    326 -----TTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCC 380
               |  |||||| ||||||||||||||||| |   |    ||    ||   ||||| |
Db    216 GGGGATGAGGAGGAGGAGGAGAAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAGTCTAC 275
```

```
Qy     381 ATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCC 440
           | |||| ||||   |||| ||| |  |  | |||| | |||||||| || || ||||||
Db     276 TTAAATATAAGCCTCCCTCCTCACACGCCGCCTCCTCCCTTCATCTCCGAGGACGCCTCC 335

Qy     441 GGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTT 500
           |||||| |||||||| |||||||||| |||||  | || || ||||||||    |||||
Db     336 GGATATCTGACCAGCCCCTGGCTGACGCTCTTCATGCCCTCCGTGTACACGATTGTGTTC 395

Qy     501 GTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
           | |||||||| || || ||| || |||||||||  |||||| || ||
Db     396 ATTGTCAGCCTTCCTCTGAACGTCCTGGCCATCGCAGTGTTCGTCTTG 443
```

RESULT 10
AA882290

LOCUS	AA882290 688 bp mRNA linear EST 26-MAR-1998
DEFINITION	vx43b01.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone IMAGE:1277929 5' similar to gb:M62424 THROMBIN RECEPTOR PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION	AA882290
VERSION	AA882290.1 GI:2991401
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 688)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:669729 Seq primer: -28m13 rev1 ET from Amersham High quality sequence stop: 314.

FEATURES	Location/Qualifiers
source	1. .688 /organism="Mus musculus" /strain="C57BL/6 x CBA" /db_xref="taxon:10090" /clone="IMAGE:1277929" /clone_lib="Stratagene mouse lung 937302" /sex="female" /tissue_type="lung" /dev_stage="6-8 month old" /lab_host="SOLR (kanamycin resistant)" /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. 6-8 month old female lung and 1.5 year old male lung were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"

BASE COUNT 126 a 212 c 188 g 162 t
ORIGIN

Query Match 32.1%; Score 175.8; DB 9; Length 688;
Best Local Similarity 65.6%; Pred. No. 8e-34;
Matches 279; Conservative 0; Mismatches 132; Indels 14; Gaps 1;

```
Qy     138 CAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGC 197
           | ||  ||||   ||||  ||   |  ||| | | |||  ||||| | || |||  | ||
Db       1 CGGCCGCGCGGGCAGCCTTGGGACAATGGGGCCCCGCGCTTGCTGATCGTCGCCCTCGGC 60

Qy     198 TTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAA 257
           |||| |||||||  ||  || ||||| |||||  ||      ||  ||||||||||  |
Db      61 CTCAGCCTGTGCGATCCCTTGCTGTCTTCCCGCGTCCCTATGAGCCAGCCAGAATCAGAG 120

Qy     258 GCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATAT 317
           ||| |||| ||  |  | ||||| |||||  |||| ||||| |||| ||| ||
Db     121 AGGACAGATGCTACGGTGAACCCCCGCTCATTCTTTCTAAGGAATCCCAGTGAAAATACA 180

Qy     318 GAACCATT--------------TTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTG 363
           | |               |  |||||| |||||||||||||||||  |   |    |
Db     181 TTTGAACTGGTCCCCCTGGGGATGAGGAGGAGGAGGAGAAAAATGAAAGCGTCCTGCTGG 240

Qy     364 AATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCA 423
           |    ||   ||||| | | |||| ||||   |||| ||| |  |  | |||| | ||||
Db     241 AGGGTAGGGCAGTCTACTTAAATATAAGCCTCCCTCCTCACACGCCGCCTCCTCCCTTCA 300

Qy     424 TCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTG 483
           |||| || || |||||||||||| |||||||| |||||||||  |||||  | || || |
Db     301 TCTCCGAGGACGCCTCCGGATATCTGACCAGCCCCTGGCTGACGCTCTTCATGCCCTCCG 360

Qy     484 TGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCA 543
           |||||||    |||||  | |||||||| || || ||| || |||||||||  ||||||
Db     361 TGTACACGATTGTGTTCATTGTCAGCCTTCCTCTGAACGTCCTGGCCATCGCAGTGTTCG 420

Qy     544 TCCTG 548
           || ||
Db     421 TCTTG 425
```

RESULT 11
BE911999

LOCUS	BE911999 672 bp mRNA linear EST 29-SEP-2000
DEFINITION	601662765F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962777 5', mRNA sequence.
ACCESSION	BE911999
VERSION	BE911999.1 GI:10408985
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 672)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM9130 row: a column: 18 High quality sequence stop: 626.

FEATURES	Location/Qualifiers
source	1. .672 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:3962777" /clone_lib="NCI_CGAP_Mam1" /tissue_type="tumor, biopsy sample" /dev_stage="10 months, virgin" /lab_host="DH10B" /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 118 a 211 c 188 g 155 t
ORIGIN

```
  Query Match             30.2%;  Score 165.6;  DB 12;  Length 672;
  Best Local Similarity   67.6%;  Pred. No. 3e-31;
  Matches  284;  Conservative    0;  Mismatches  119;  Indels   17;  Gaps    3;

Qy     144 CGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCCGCCTGCTTCAGT 203
           ||||   |||| |||||||||||||| |||||  |||||  | || |||  | || ||||
Db      19 CGCGGGCAGCCTGGGACAATGGGGCCCCGGCGCTTGCTGATCGTCGCCCTCGGC-TCAGC  77

Qy     204 CTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACA 263
           |||||||| ||  || ||||| |||||  ||      ||  |||||||||| |    |||
Db      78 CTGTGCGGTCCCTTGCTGTCTTCCCGCGTCCCTATGAGCCAGCCAGAATCAGAGAGGACA 137

Qy     264 AATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCA 323
           ||||  ||  |  | |||||  |||||  ||||  |||| |||| ||| ||         |
Db     138 GATGCTACGGTGAACCCCCGCTCATTCTTTCTATGGAATCCCAGTGAAAATACATTTGAA 197

Qy     324 TT--------------TTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATAC 368
           ||               |  ||||||  |||||||||||||||||  |   |    ||
Db     198 CTGGTCCCCCTGGGGGATGAGGAGGAGGAGGAGAAAAATGAAAGCGTCCTGCTGGAGGGT 257

Qy     369 AGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCA 428
           ||   |||||  | | |||| ||||   |||| ||| |  |  | |||| | ||||||||
Db     258 AGGGCAGTCTACTTAAATATAAGCCTCCCTCCTCACACGCCGC-TCCTCCCTTCATCTCC 316

Qy     429 GAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTAC 488
           || || |||||||||||| |||||||| |||||||||| |||||  | || || ||||||
Db     317 GAGGACGCCTCCGGATATCTGACCAGCCCCTGGCTGACGCTCTTCATGCCCTCCGTGTAC 376

Qy     489 ACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
           ||    |||||  | |||||||| || || ||| || |||||||||  |||||| || ||
Db     377 ACGATTGTGTTCATTGTCAGCCTTCCTCTGAACGTCCTGGCCATCGCAGTGTTCGTCTTG 436
```

```
RESULT 12
BI666195
LOCUS       BI666195                763 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603287191F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5321110 5',
            mRNA sequence.
ACCESSION   BI666195
VERSION     BI666195.1  GI:15580428
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 763)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I:M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11814  row: f  column: 23
            High quality sequence stop: 756.
FEATURES             Location/Qualifiers
     source          1. .763
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5321110"
                     /clone_lib="NCI_CGAP_Mam6"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
            Library constructed by Life Technologies. Investigator
            providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      138 a    238 c    216 g    170 t      1 others
ORIGIN
```

```
  Query Match             29.6%;  Score 162.2;  DB 13;  Length 763;
  Best Local Similarity   65.3%;  Pred. No. 2.2e-30;
  Matches  292;  Conservative    0;  Mismatches  138;  Indels   17;  Gaps    3;

Qy     117 AGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGG 176
           ||  || |  | |  ||  |  |||  |    |  |||||||||||||| |||||
Db      50 AGTAGGCGACGGCGGGCGCCGGGCCGCGCGGGCAGCCTTGGGACAATGGGGCCCCGGCGC 109

Qy     177 CTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGG 236
           |||||  | || |||  | || |||| |||||||| ||  || |||||  |||||  ||
Db     110 TTGCTGATCGTCGCCCTCGGC-TCAGCCTGTGCGGTCCCTTGCTGTCTTCCCGCGTCCTA 168

Qy     237 GCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTC 296
              || |||||||||| |    ||| |||| ||  |  | |||||  |||||  ||||
Db     169 TGAGCCA-GCCAGAATCAGAGAGGACAGATGCTACGGTGAACCCCCGCTCATTCTTTCTA 227

Qy     297 AGGAACCCCAATGATAAATATGAACCATT---------------TTGGGAGGATGAGGAG 341
           |||||  |||| |||  ||          | |               |  ||||||  ||||||
Db     228 AGGAATCCCAGTGAAAATACATTTGAACTGGTCCCCCTGGGGGATGAGGAGGAGGAGGAG 287

Qy     342 AAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTT 401
           ||||||||||| |   |    ||    ||   |||||  | | |||| ||||   |||| |
Db     288 AAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAGTCTACTTAAATATAAGCCTCCCTCCT 347

Qy     402 CAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGG 461
           || |  |  | |||| | |||||||| || || ||||||||||| |||||||| |||||
Db     348 CACACGCCGCCTCCTCCCTTCATCTCCGAGGACGCCTCCGGATATCTGACCAGCCCCTGG 407

Qy     462 CTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAAC 521
           |||||  |||||  | || || ||||||||    |||||  | |||||||  || || |||
Db     408 CTGACGCTCTTCATGCCCTCCGTGTACACGATTGTGTTCATTGTCAGCCTTCCTCTGAAC 467

Qy     522 ATCATGGCCATCGTTGTGTTCATCCTG 548
           || |||||||||  ||||||  || ||
Db     468 GTCCTGGCCATCGCAGTGTTCGTCTTG 494
```

```
RESULT 13
BB862683
LOCUS       BB862683                467 bp    mRNA    linear   EST 26-NOV-2001
DEFINITION  BB862683 RIKEN full-length enriched, RCB-0035 WEHI-3 cDNA Mus
            musculus cDNA clone G430029F09 5', mRNA sequence.
ACCESSION   BB862683
VERSION     BB862683.1  GI:17104137
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
```

```
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G430029F09"
                     /clone_lib="RIKEN full-length enriched, RCB-0035 WEHI-3
                     cDNA"
                     /cell_line="RCB-0035 WEHI-3"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                     ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_type=stroma cell,
                     cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                     cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                     ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                     BC3H1)"
BASE COUNT       91 a    149 c    130 g     97 t
ORIGIN

  Query Match             29.1%;  Score 159.2;  DB 10;  Length 467;
  Best Local Similarity   64.1%;  Pred. No. 1.1e-29;
  Matches  264;  Conservative    0;  Mismatches  133;  Indels   15;  Gaps    1;

Qy     117 AGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGG 176
           ||  || |  | |   ||  |   |||   |     |  ||||||||||||| |||||
Db      46 AGTAGGCGACGGCGGGCGCCGGGCCGCGCGGGCAGCCTTGGGACAATGGGGCCCCGGCGC 105

Qy     177 CTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGG 236
           |||||  |  ||  |||  | ||  |||| ||||||||  ||  || |||||  |||||  ||
Db     106 TTGCTGATCGTCGCCCTCGGCCTCAGCCTGTGCGGTCCCTTGCTGTCTTCCCGCGTCCCT 165

Qy     237 GCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTC 296
             ||  |||||||||| |    |||  |||| ||  |  | |||||  |||||  ||||
Db     166 ATGAGCCAGCCAGAATCAGAGAGGACAGATGCTACGGTGAACCCCCGCTCATTCTTTCTA 225

Qy     297 AGGAACCCCAATGATAAATATGAACCATT--------------TTGGGAGGATGAGGAG 341
           |||||  ||||  |||  ||         | |                |  |||||| ||||||
Db     226 AGGAATCCCAGTGAAAATACATTTGAACTGGTCCCCCTGGGGGATGAGGAGGAGGAGGAG 285

Qy     342 AAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTT 401
           ||||||||||| |   |    ||    ||   ||||| | | ||||  ||||   |||| |
Db     286 AAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAGTCTACTTAAATATAAGCCTCCCTCCT 345

Qy     402 CAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGG 461
           || |  |  | |||| | |||||||| || || |||||||||||| |||||||| |||||
Db     346 CACACGCCGCCTCCTCCCTTCATCTCCGAGGACGCCTCCGGATATCTGACCAGCCCCTGG 405

Qy     462 CTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCC 513
           ||||| |||||  | || || ||||||||    |||||  | || ||||| |
Db     406 CTGACGCTCTTCATGCCCTCCGTGTACACGATTGTGTTCATTGTTAGCCTTC 457


RESULT 14
BB852109
LOCUS       BB852109                466 bp    mRNA    linear   EST 26-NOV-2001
DEFINITION  BB852109 RIKEN full-length enriched, B16 F10Y cells Mus musculus
            cDNA clone G370005M18 5', mRNA sequence.
ACCESSION   BB852109
VERSION     BB852109.1  GI:17093563
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="G370005M18"
                     /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
                     /cell_type="B16 F10Y cells"
                     /note="pooled tissues ; (tissue_type=cerebellum,
                     dev_stage=16 days neonate, sex=mixed),
```

(tissue_type=cerebellum, dev_stage=0 day neonate, sex=mixed), (tissue_type=hippocampus, dev_stage=adult, sex=male), (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days embryo, sex=mixed)"

BASE COUNT 90 a 145 c 130 g 101 t
ORIGIN

Query Match 28.8%; Score 157.6; DB 10; Length 466;
Best Local Similarity 63.5%; Pred. No. 2.8e-29;
Matches 283; Conservative 0; Mismatches 144; Indels 19; Gaps 2;

```
Qy    117 AGGCGGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGG 176
          ||  || |  | |    ||  | | |||   |     |  ||||||||||||||| |||||
Db     25 AGTAGGCGACGGCGGGCGCCGGACCGCGCGGGCAGCCTTGGGACAATGGGGCCCCGGCGC 84

Qy    177 CTGCTGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGG 236
          ||||| | || |||  | || |||| |||||||| ||  || ||||| |||||  ||
Db     85 TTGCTGATCGTCGCCCTCGGCCTCAGCCTGTGCGGTCCCTTGCTGTCTTCCCGCGTCCCT 144

Qy    237 GCCCGCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTC 296
          ||  |||||||||| |    ||| |||| ||  |  | |||| |||||  ||||
Db    145 ATGAGCCAGCCAGAATCAGAGAGGACAGATGCTACGGTGAACCCCCGCTCATTCTTTCTA 204

Qy    297 AGGAACCCCAATGATAAATATGAACCATT--------------TTGGGAGGATGAGGAG 341
          ||||| |||| ||| ||         | |              |  |||||| ||||||
Db    205 AGGAATCCCAGTGAAAATACATTTGAACTGGTCCCCCTGGGGGATGAGGAGGAGGAGGAG 264

Qy    342 AAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTT 401
          ||||||||||| |   |    ||    ||   ||||| | | |||| ||||   |||| |
Db    265 AAAAATGAAAGCGTCCTGCTGGAGGGTAGGGCAGTCTACTTAAATATAAGCCTCCCTCCT 324

Qy    402 CAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGG 461
          || |  |  | |||| | |||||||| || || |||||||||||| |||||||| |||||
Db    325 CACACGCCGCCTCCTCCCTTCATCTCCGAGGACGCCTCCGGATATCTGACCAGCCCCTGG 384

Qy    462 CTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAAC 521
          ||||| |||||  | ||   |||    |  |       |  | |||||||| || || |||
Db    385 CTGACGCTCTTCATGCCTCGTGT----AAGGATTGTGTCATTGTCAGCCTTCCTCTGAAC 440

Qy    522 ATCATGGCCATCGTTGTGTTCATCCT 547
            || |||||||||  ||||||  || |
Db    441 GTCCTGGCCATCGCAGTGTTCGTCTT 466
```

RESULT 15
BQ926183
LOCUS BQ926183 1429 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8863932 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6312003 5', mRNA sequence.
ACCESSION BQ926183
VERSION BQ926183.1 GI:22341203
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1429)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13736 row: n column: 04
High quality sequence start: 62
High quality sequence stop: 569.

FEATURES Location/Qualifiers
source 1. .1429
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6312003"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

BASE COUNT 309 a 428 c 417 g 266 t 9 others
ORIGIN

Query Match 27.7%; Score 152; DB 14; Length 1429;
Best Local Similarity 66.7%; Pred. No. 9.7e-28;
Matches 240; Conservative 0; Mismatches 105; Indels 15; Gaps 1;

```
Qy    204 CTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAATCAAAAGCAACA 263
          |||||||| ||  || ||||| |||||  ||      ||  |||||||||| |    |||
Db     88 CTGTGCGGTCCCTTGCTGTCTTCCCGCGTCCCTATGAGCCAGCCAGAATCAGAGAGGACA 147

Qy    264 AATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCA 323
          ||||  ||  |  | |||||  |||||  ||||  |||||  ||||  |||  ||         |
Db    148 GATGCTACGGTGAACCCCCGCTCATTCTTTCTAAGGAATCCCAGTGAAAATACATTTGAA 207

Qy    324 TT--------------TTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATAC 368
           |              |  ||||||  ||||||||||||||||  |   |    ||
Db    208 CTGGTCCCCCTGGGGGATGAGGAGGAGGAGGAGAAAAATGAAAGCGTCCTGCTGGAGGGT 267

Qy    369 AGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCA 428
          ||   ||||| | | |||| ||||   |||| ||| |   |  | |||| | ||||||||
Db    268 AGGGCAGTCTACTTAAATATAAGCCTCCCTCCTCACACGCCGCCTCCTCCCTTCATCTCC 327

Qy    429 GAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTAC 488
          || || |||||||||||| |||||||| |||||||||| |||||  | || || ||||||
Db    328 GAGGACGCCTCCGGATATCTGACCAGCCCCTGGCTGACGCTCTTCATGCCCTCCGTGTAC 387

Qy    489 ACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
          ||    |||||  | |||||||| || || ||| || |||||||||  |||||| || ||
Db    388 ACGATTGTGTTCATTGTCAGCCTTCCTCTGAACGTCCTGGCCATCGCAGTGTTCGTCTTG 447
```

Search completed: December 9, 2002, 05:49:59
Job time : 2456 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 04:05:26 ; Search time 3119 Seconds
(without alignments)
5113.286 Million cell updates/sec

Title: US-09-744-679-7
Perfect score: 548
Sequence: 1 cgccgagggtcgcttggacc.....ccatcgttgtgttcatcctg 548

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		548	100.0	3472	9	HUMTHRR	M62424 Human throm
2		548	100.0	3479	9	BC002464	BC002464 Homo sapi
3		548	100.0	3480	6	AR025392	AR025392 Sequence
4		548	100.0	3480	6	AR027730	AR027730 Sequence
5		548	100.0	3480	6	AR065265	AR065265 Sequence
6		548	100.0	3480	6	AR137610	AR137610 Sequence
7		548	100.0	3480	6	I74660	I74660 Sequence 21
8		441.6	80.6	3764	9	AF028727	AF028727 Papio ham
9		306.2	55.9	70679	2	AC026151	AC026151 Homo sapi
10		302	55.1	24771	9	AF391809	AF391809 Homo sapi
11		302	55.1	110965	2	AC020927	AC020927 Homo sapi
12		302	55.1	143335	9	AC025188	AC025188 Homo sapi
13		282.4	51.5	1764	6	I49726	I49726 Sequence 12
14		249	45.4	88881	2	AC008933	AC008933 Homo sapi
15		246	44.9	2895	9	HSU63331	U63331 Human throm
16		229	41.8	2878	9	AF054633	AF054633 Homo sapi
17		188.2	34.3	2855	10	CLTHRREC	X61958 C.longicaud
18		184.2	33.6	1312	10	MUSTHRRCT	L03529 Mus musculu
19		183	33.4	3313	10	BC031516	BC031516 Mus muscu
20		171.2	31.2	3418	10	RATTRGPC	M81642 Rat G-prote
21		161	29.4	5457	9	HSU36755	U36755 Human throm
22	C	148.2	27.0	6436	6	AX345583	AX345583 Sequence
23		146.6	26.8	6436	6	AX345582	AX345582 Sequence
24	C	137.6	25.1	151166	2	AC130634	AC130634 Rattus no
25		133	24.3	5620	10	MMTHREC02	U36757 Mus musculu
26		133	24.3	199227	2	AC110876	AC110876 Mus muscu
27		130.4	23.8	1164	10	CLU34047	U34047 Cricetulus
28		53	9.7	275	10	MMCF2R01	U55075 Mus musculu
29		53	9.7	4605	10	MMTHREC01	U36756 Mus musculu
30	C	53	9.7	235097	2	AC110877	AC110877 Mus muscu
31		52.6	9.6	1680	5	XLU09632	U09632 Xenopus lae
32		48.8	8.9	47981	1	AF263245	AF263245 Micromono
33		48.8	8.9	47981	6	AX112026	AX112026 Sequence
34	C	47.2	8.6	195670	2	AC115449	AC115449 Rattus no
35	C	47	8.6	2464	10	AY034062	AY034062 Mus muscu
36	C	47	8.6	69747	2	AC124340	AC124340 Mus muscu
37	C	47	8.6	193947	10	AL606985	AL606985 Mouse DNA
38	C	47	8.6	195554	2	AL714009	AL714009 Mus muscu
39	C	46.4	8.5	86237	2	AC111966	AC111966 Rattus no
40		46	8.4	91856	2	AC123018	AC123018 Rattus no
41		46	8.4	118155	2	AC130148	AC130148 Rattus no
42	C	45.2	8.2	157861	2	AC126902	AC126902 Rattus no
43	C	45	8.2	134725	2	AC114389	AC114389 Rattus no
44		45	8.2	175344	2	AC097596	AC097596 Rattus no
45		44.8	8.2	159822	2	AC118304	AC118304 Rattus no

ALIGNMENTS

RESULT 1
HUMTHRR
LOCUS HUMTHRR 3472 bp mRNA linear PRI 03-AUG-1993
DEFINITION Human thrombin receptor mRNA, complete cds.
ACCESSION M62424
VERSION M62424.1 GI:339676
KEYWORDS thrombin receptor.
SOURCE Human DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3472)
AUTHORS Vu,T.K., Hung,D.T., Wheaton,V.I. and Coughlin,S.R.
TITLE Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation

```
  JOURNAL   Cell 64 (6), 1057-1068 (1991)
  MEDLINE   91168254
   PUBMED   1672265
FEATURES             Location/Qualifiers
     source          1. .3472
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             225. .1502
                     /codon_start=1
                     /product="thrombin receptor"
                     /protein_id="AAA36743.1"
                     /db_xref="GI:339677"
                     /translation="MGPRRLLLVAACFSLCGPLLSARTRARRPESKATNATLDPRSFL
                     LRNPNDKYEPFWEDEEKNESGLTEYRLVSINKSSPLQKQLPAFISEDASGYLTSSWLT
                     LFVPSVYTGVFVVSLPLNIMAIVVFILKMKVKKPAVVYMLHLATADVLFVSVLPFKIS
                     YYFSGSDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTLGR
                     ASFTCLAIWALAIAGVVPLVLKEQTIQVPGLNITTCHDVLNETLLEGYYAYYFSAFSA
                     VFFFVPLIISTVCYVSIIRCLSSSAVANRSKKSRALFLSAAVFCIFIICFGPTNVLLI
                     AHYSFLSHTSTTEAAYFAYLLCVCVSSISSCIDPLIYYYASSECQRYVYSILCCKESS
                     DPSSYNSSGQLMASKMDTCSSNLNNSIYKKLLT"
• BASE COUNT      933 a    817 c    785 g    937 t
  ORIGIN
```

Query Match 100.0%; Score 548; DB 9; Length 3472;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603

Qy     541 TCATCCTG 548
           ||||||||
Db     604 TCATCCTG 611
```

```
RESULT 2
BC002464
LOCUS       BC002464                3479 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, coagulation factor II (thrombin) receptor, clone
            MGC:1197 IMAGE:3343051, mRNA, complete cds.
ACCESSION   BC002464
VERSION     BC002464.1  GI:12803296
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3479)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site:       http://www.nisc.nih.gov/
            Contact:        nisc_mgc@nhgri.nih.gov
            Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
            Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 5 Row: p Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6031164.
FEATURES             Location/Qualifiers
     source          1. .3479
                     /organism="Homo sapiens"
                     /db_xref="LocusID:2149"
                     /db_xref="taxon:9606"
                     /clone="MGC:1197 IMAGE:3343051"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_16"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             190. .1467
                     /codon_start=1
                     /product="coagulation factor II (thrombin) receptor"
                     /protein_id="AAH02464.1"
                     /db_xref="GI:12803297"
                     /translation="MGPRRLLLVAACFSLCGPLLSARTRARRPESKATNATLDPRSFL
                     LRNPNDKYEPFWEDEEKNESGLTEYRLVSINKSSPLQKQLPAFISEDASGYLTSSWLT
                     LFVPSVYTGVFVVSLPLNIMAIVVFILKMKVKKPAVVYMLHLATADVLFVSVLPFKIS
                     YYFSGSDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTLGR
                     ASFTCLAIWALAIAGVVPLLLKEQTIQVPGLNITTCHDVLNETLLEGYYAYYFSAFSA
                     VFFFVPLIISTVCYVSIIRCLSSSAVANRSKKSRALFLSAAVFCIFIICFGPTNVLLI
                     VHYSFLSHTSTTEAAYFAYLLCVCVSSISCCIDPLIYYYASSECQRYVYSILCCKESS
                     DPSSYNSSGQLMASKMDTCSSNLNNSIYKKLLT"
BASE COUNT      969 a    792 c    778 g    940 t
ORIGIN
```

Query Match 100.0%; Score 548; DB 9; Length 3479;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 88

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
```

```
Db      89 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 148

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
Db     149 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 208

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
Db     209 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 268

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
Db     269 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 328

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
Db     329 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 388

Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
Db     389 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 448

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
Db     449 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 508

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
Db     509 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 568

Qy     541 TCATCCTG 548
Db     569 TCATCCTG 576
```

```
RESULT 3
AR025392
LOCUS       AR025392               3480 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 219 from patent US 5798248.
ACCESSION   AR025392
VERSION     AR025392.1  GI:3978020
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3480)
  AUTHORS   Coughlin,S.R. and Scarborough,R.M.
  TITLE     Recombinant mutant thrombin receptor and related pharmaceuticals
  JOURNAL   Patent: US 5798248-A 219 25-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .3480
                     /organism="unknown"
BASE COUNT      941 a    817 c    785 g    937 t
ORIGIN

  Query Match             100.0%;  Score 548;  DB 6;  Length 3480;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-112;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
Db      64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
Db     124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
Db     184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
```

```
Db     244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
Db     304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
Db     364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
Db     424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
Db     484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
Db     544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603

Qy     541 TCATCCTG 548
Db     604 TCATCCTG 611
```

```
RESULT 4
AR027730
LOCUS       AR027730               3480 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 219 from patent US 5856448.
ACCESSION   AR027730
VERSION     AR027730.1  GI:5938550
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3480)
  AUTHORS   Coughlin,S.R.
  TITLE     Antibodies specifically reactive with thrombin receptor and its
            components
  JOURNAL   Patent: US 5856448-A 219 05-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .3480
                     /organism="unknown"
BASE COUNT      941 a    817 c    785 g    937 t
ORIGIN

  Query Match             100.0%;  Score 548;  DB 6;  Length 3480;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-112;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
Db      64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
Db     124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
Db     184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
Db     244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
Db     304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363
```

```
Qy    301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy    361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy    421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy    481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603

Qy    541 TCATCCTG 548
          ||||||||
Db    604 TCATCCTG 611
```

```
RESULT 5
AR065265
LOCUS       AR065265               3480 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 219 from patent US 5849507.
ACCESSION   AR065265
VERSION     AR065265.1  GI:5995481
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3480)
  AUTHORS   Coughlin,S.R.
  TITLE     Methods to diagnose thrombosis by measuring activation peptide
  JOURNAL   Patent: US 5849507-A 219 15-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .3480
                     /organism="unknown"
BASE COUNT      941 a    817 c    785 g    937 t
ORIGIN

  Query Match           100.0%;  Score 548;  DB 6;  Length 3480;
  Best Local Similarity 100.0%;  Pred. No. 3.3e-112;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy     61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy    121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy    181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy    241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363

Qy    301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy    361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy    421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy    481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603

Qy    541 TCATCCTG 548
          ||||||||
Db    604 TCATCCTG 611
```

```
RESULT 6
AR137610
LOCUS       AR137610               3480 bp    DNA     linear   PAT 16-JUN-2001
DEFINITION  Sequence 219 from patent US 6197541.
ACCESSION   AR137610
VERSION     AR137610.1  GI:14479119
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3480)
  AUTHORS   Coughlin,S.R.
  TITLE     Recombinant thrombin receptors and assays using them
  JOURNAL   Patent: US 6197541-A 219 06-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .3480
                     /organism="unknown"
BASE COUNT      941 a    817 c    785 g    937 t
ORIGIN

  Query Match           100.0%;  Score 548;  DB 6;  Length 3480;
  Best Local Similarity 100.0%;  Pred. No. 3.3e-112;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy     61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy    121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy    181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy    241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363

Qy    301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy    361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy    421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy    481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603
```

```
Qy     541 TCATCCTG 548
           ||||||||
Db     604 TCATCCTG 611


RESULT 7
I74660
LOCUS       I74660                  3480 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 219 from patent US 5688768.
ACCESSION   I74660
VERSION     I74660.1  GI:3010801
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3480)
  AUTHORS   Coughlin,S.R. and Scarborough,R.M.
  TITLE     Recombinant thrombin receptor and related pharmaceuticals
  JOURNAL   Patent: US 5688768-A 219 18-NOV-1997;
FEATURES             Location/Qualifiers
     source          1. .3480
                     /organism="unknown"
BASE COUNT      941 a    817 c    785 g    937 t
ORIGIN

  Query Match           100.0%;  Score 548;  DB 6;  Length 3480;
  Best Local Similarity  100.0%;  Pred. No. 3.3e-112;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 123

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 243

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 303

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 363

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 423

Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 483

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     484 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 543

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603

Qy     541 TCATCCTG 548
           ||||||||
Db     604 TCATCCTG 611


RESULT 8
AF028727
LOCUS       AF028727                3764 bp    mRNA    linear   PRI 20-NOV-2000
DEFINITION  Papio hamadryas thrombin receptor (bthr12) mRNA, complete cds.
ACCESSION   AF028727
VERSION     AF028727.1  GI:2605876
KEYWORDS    .
SOURCE      Papio hamadryas.
  ORGANISM  Papio hamadryas
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 3764)
  AUTHORS   Hayzer,D.J., Shoji,M. and Hanson,S.R.
  TITLE     cDNAs encoding the baboon thrombin receptor indicate a primate
            transcription start site upstream of putative sites reported for
            the human gene
  JOURNAL   Thromb. Res. 98 (2), 195-201 (2000)
  MEDLINE   20180190
   PUBMED   10713321
REFERENCE   2  (bases 1 to 3764)
  AUTHORS   Shoji,M., Hayzer,D.J. and Hanson,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1997) Hematologic Diseases Division, Building 1,
            CDC/NCID/DHA, Rm. 1310, Mail Stop D-02, 1600 Clifton Road, Atlanta,
            GA 30333, USA
FEATURES             Location/Qualifiers
     source          1. .3764
                     /organism="Papio hamadryas"
                     /db_xref="taxon:9557"
     gene            1. .3764
                     /gene="bthr12"
     CDS             532. .1809
                     /gene="bthr12"
                     /codon_start=1
                     /product="thrombin receptor"
                     /protein_id="AAB84191.1"
                     /db_xref="GI:2605877"
                     /translation="MGPRRLLLVAACLCLCGPLLSARTRARRPASKATNATLDPRSFL
                     LRNPNDKYEPFWEDEEKNESGLTEYRLVSINKSSPLQKPLPAFISEDASGYLTSSWLT
                     LFVPSVYTGVFVVSLPVNIMAIVVFILKMKVKKPAVVYMLHLATADVLFVSVLPFKIS
                     YYLSGSDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTLGR
                     ASFTCLAIWALAIAGVVPLLLKEQTIQVPGLNITTCHDVLNETLLEGYYAYYFSAFSA
                     VFFFVPLIISTVCYVSIIRCLSSSTVANRSKKSRALFLSAAVFCIFIICFGPTNILLI
                     AHYSFLSHTSTTEAAYFAYLLCVCVSSISCCIDPLIYYYASSECQRYVYSILCCKESS
                     DPSSSNSSGQLMASKMDTCSSNLNNSIYKKLLT"
BASE COUNT      992 a    868 c    863 g   1041 t
ORIGIN

  Query Match           80.6%;  Score 441.6;  DB 9;  Length 3764;
  Best Local Similarity  91.8%;  Pred. No. 2e-88;
  Matches  503;  Conservative    0;  Mismatches   34;  Indels   11;  Gaps    3;

Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
           ||||| |||||||| |||||||| |||| |||| ||          || ||||||||||||
Db     382 CGCCGGGGGTCGCTCGGACCCTGCTCTTGCCCGCGG---------GCGCTGCCTGCCGCG 432

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
           | |||||||| ||||||| ||||||||  | |||||||||| | ||||||||||| ||||
Db     433 ACGACCGGCTGCCCGACCAGCAGAAGTTGGTAGAGAGGGTGGA-CGGAGCAGCCCAAGGC 491

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
           |||| | ||||| |||||||||| | || || ||||||||||||||||||||||||||||
Db     492 GGGG-AACCTCCTGGAGCAGCGCAGTGCGGATCCCGGGACAATGGGGCCGCGGCGGCTGC 550

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
           |||||||||||||||| || |||||||||||||||||||| || ||||||||||||||||
Db     551 TGCTGGTGGCCGCCTGCCTCTGTCTGTGCGGCCCGCTGTTGTCCGCACGCACCCGGGCCC 610

Qy     241 GCAGGCCAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGA 300
           |||||||||  |||||||||||||||||||| ||||||||||||||||||||||||||||
Db     611 GCAGGCCAGCGTCAAAAGCAACAAATGCCACGTTAGATCCCCGGTCATTTCTTCTCAGGA 670

Qy     301 ACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 360
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     671 ATCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAA 730
```

```
Qy     361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCAT 420
           ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |
Db     731 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACCACTTCCTGCGT 790

Qy     421 TCATCTCAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 480
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     791 TCATCTCAGAAGATGCCTCCGGATATCTGACCAGCTCCTGGCTGACACTCTTTGTCCCAT 850

Qy     481 CTGTGTACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
           |||||||||||| ||||||||||||||||||  ||||||||||||||||||||||||||
Db     851 CTGTGTACACCGGGGTGTTTGTAGTCAGCCTCCCGGTAAACATCATGGCCATCGTTGTGT 910

Qy     541 TCATCCTG 548
           ||||||||
Db     911 TCATCCTG 918
```

RESULT 9
AC026151
LOCUS AC026151 70679 bp DNA linear HTG 19-MAR-2000
DEFINITION Homo sapiens clone RP11-329A20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC026151
VERSION AC026151.1 GI:7264221
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 70679)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-329A20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 70679)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L8625
Center clone name: 329_A_20
----------.
* NOTE: This record contains 77 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 849: contig of 849 bp in length
* 850 949: gap of 100 bp
* 950 1768: contig of 819 bp in length
* 1769 1868: gap of 100 bp
* 1869 2679: contig of 811 bp in length
* 2680 2779: gap of 100 bp
* 2780 3583: contig of 804 bp in length
* 3584 3683: gap of 100 bp
* 3684 4465: contig of 782 bp in length
* 4466 4565: gap of 100 bp
* 4566 5401: contig of 836 bp in length
* 5402 5501: gap of 100 bp
* 5502 6340: contig of 839 bp in length
* 6341 6440: gap of 100 bp
* 6441 7270: contig of 830 bp in length
* 7271 7370: gap of 100 bp
* 7371 8217: contig of 847 bp in length
* 8218 8317: gap of 100 bp
* 8318 9141: contig of 824 bp in length
* 9142 9241: gap of 100 bp
* 9242 10057: contig of 816 bp in length
* 10058 10157: gap of 100 bp
* 10158 10986: contig of 829 bp in length
* 10987 11086: gap of 100 bp
* 11087 11911: contig of 825 bp in length
* 11912 12011: gap of 100 bp
* 12012 12788: contig of 777 bp in length
* 12789 12888: gap of 100 bp
* 12889 13702: contig of 814 bp in length
* 13703 13802: gap of 100 bp
* 13803 14634: contig of 832 bp in length
* 14635 14734: gap of 100 bp
* 14735 15571: contig of 837 bp in length
* 15572 15671: gap of 100 bp
* 15672 16504: contig of 833 bp in length
* 16505 16604: gap of 100 bp
* 16605 17393: contig of 789 bp in length
* 17394 17493: gap of 100 bp
* 17494 18327: contig of 834 bp in length
* 18328 18427: gap of 100 bp
* 18428 19259: contig of 832 bp in length
* 19260 19359: gap of 100 bp
* 19360 20208: contig of 849 bp in length
* 20209 20308: gap of 100 bp
* 20309 21141: contig of 833 bp in length
* 21142 21241: gap of 100 bp
* 21242 22060: contig of 819 bp in length
* 22061 22160: gap of 100 bp
* 22161 22966: contig of 806 bp in length
* 22967 23066: gap of 100 bp
* 23067 23869: contig of 803 bp in length
* 23870 23969: gap of 100 bp
* 23970 24798: contig of 829 bp in length
* 24799 24898: gap of 100 bp
* 24899 25702: contig of 804 bp in length
* 25703 25802: gap of 100 bp
* 25803 26567: contig of 765 bp in length
* 26568 26667: gap of 100 bp
* 26668 27483: contig of 816 bp in length
* 27484 27583: gap of 100 bp
* 27584 28412: contig of 829 bp in length
* 28413 28512: gap of 100 bp
* 28513 29344: contig of 832 bp in length
* 29345 29444: gap of 100 bp
* 29445 30271: contig of 827 bp in length

```
*    30272 30371: gap of      100 bp
*    30372    31143: contig of 772 bp in length
*    31144 31243: gap of      100 bp
*    31244    32053: contig of 810 bp in length
*    32054 32153: gap of      100 bp
*    32154    32962: contig of 809 bp in length
*    32963 33062: gap of      100 bp
*    33063    33883: contig of 821 bp in length
*    33884 33983: gap of      100 bp
*    33984    34807: contig of 824 bp in length
*    34808 34907: gap of      100 bp
*    34908    35712: contig of 805 bp in length
*    35713 35812: gap of      100 bp
*    35813    36643: contig of 831 bp in length
*    36644 36743: gap of      100 bp
*    36744    37572: contig of 829 bp in length
*    37573 37672: gap of      100 bp
*    37673    38499: contig of 827 bp in length
*    38500 38599: gap of      100 bp
*    38600    39425: contig of 826 bp in length
*    39426 39525: gap of      100 bp
*    39526    40305: contig of 780 bp in length
*    40306 40405: gap of      100 bp
*    40406    41233: contig of 828 bp in length
*    41234 41333: gap of      100 bp
*    41334    42152: contig of 819 bp in length
*    42153 42252: gap of      100 bp
*    42253    43091: contig of 839 bp in length
*    43092 43191: gap of      100 bp
*    43192    44035: contig of 844 bp in length
*    44036 44135: gap of      100 bp
*    44136    44973: contig of 838 bp in length
*    44974 45073: gap of      100 bp
*    45074    45888: contig of 815 bp in length
*    45889 45988: gap of      100 bp
*    45989    46837: contig of 849 bp in length
*    46838 46937: gap of      100 bp
*    46938    47765: contig of 828 bp in length
*    47766 47865: gap of      100 bp
*    47866    48665: contig of 800 bp in length
*    48666 48765: gap of      100 bp
*    48766    49576: contig of 811 bp in length
*    49577 49676: gap of      100 bp
*    49677    50500: contig of 824 bp in length
*    50501 50600: gap of      100 bp
*    50601    51437: contig of 837 bp in length
*    51438 51537: gap of      100 bp
*    51538    52369: contig of 832 bp in length
*    52370 52469: gap of      100 bp
*    52470    53296: contig of 827 bp in length
*    53297 53396: gap of      100 bp
*    53397    54202: contig of 806 bp in length
*    54203 54302: gap of      100 bp
*    54303    55128: contig of 826 bp in length
*    55129 55228: gap of      100 bp
*    55229    56078: contig of 850 bp in length
*    56079 56178: gap of      100 bp
*    56179    57000: contig of 822 bp in length
*    57001 57100: gap of      100 bp
*    57101    57897: contig of 797 bp in length
*    57898 57997: gap of      100 bp
*    57998    58820: contig of 823 bp in length
*    58821 58920: gap of      100 bp
*    58921    59718: contig of 798 bp in length
*    59719 59818: gap of      100 bp
*    59819    60609: contig of 791 bp in length
*    60610 60709: gap of      100 bp
*    60710    61529: contig of 820 bp in length
*    61530 61629: gap of      100 bp
*    61630    62460: contig of 831 bp in length
*    62461 62560: gap of      100 bp
*    62561    63354: contig of 794 bp in length
*    63355 63454: gap of      100 bp
```

```
*    63455    64251: contig of 797 bp in length

Query Match              55.9%;  Score 306.2;  DB 2;  Length 70679;
Best Local Similarity    69.6%;  Pred. No. 3.9e-58;
Matches  332;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy      72 CCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGCGGGGCAGCCTC 131
           || |||    |  |||  ||   |
Db   51413 CCAGACAGATAAGAGTATGGGAGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 51472

Qy     132 CCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGCTGCTGGTGGCC 191

Db   51473 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 51532

Qy     192 GCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCCGCAGGCCAGAA 251
             || |   ||| | ||      || |  |   |   |   || |    |||||
Db   51533 NNNNNGCACACTTGTTGCTGACCCTGCAGGTTTTTCTCTAGNAGNATCCCCCCTTCAGAA 51592

Qy     252 TCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGAT 311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   51593 TCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGAT 51652

Qy     312 AAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGA 371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   51653 AAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGA 51712

Qy     372 TTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAA 431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   51713 TTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAA 51772

Qy     432 GATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACC 491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   51773 GATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACC 51832

Qy     492 GGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   51833 GGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 51889


RESULT 10
AF391809
LOCUS       AF391809               24771 bp    DNA     linear   PRI 20-JUL-2001
DEFINITION  Homo sapiens coagulation factor II (thrombin) receptor (F2R) gene,
            complete cds.
ACCESSION   AF391809
VERSION     AF391809.2  GI:14971463
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 24771)
  AUTHORS   Rieder,M.J., Carrington,D.P., Chung,M.-W., Lee,K.L., Poel,C.L.,
            Yi,Q. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2001) Molecular Biotechnology, University of
            Washington, 1705 NE Pacific, Seattle, WA 98195, USA
REFERENCE   2  (bases 1 to 24771)
  AUTHORS   Rieder,M.J. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Molecular Biotechnology, University of
            Washington, 1705 NE Pacific, Seattle, WA 98195, USA
  REMARK    Sequence update by submitter
COMMENT     On Jul 20, 2001 this sequence version replaced gi:14583134.
            To cite this work please use:  SeattleSNPs. NHLBI Program for
            Genomic Applications, UW-FHCRC, Seattle, WA (URL:
            http://pga.mbt.washington.edu).
FEATURES             Location/Qualifiers
     source          1. .24771
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     repeat_region   55. .211
                     /rpt_family="MER104"
```

```
                    /rpt_type=dispersed
variation           234
                    /frequency="0.07"
                    /replace="C"
variation           352
                    /frequency="0.42"
                    /replace="A"
variation           636
                    /frequency="0.01"
                    /replace="A"
variation           709
                    /frequency="0.02"
                    /replace="T"
repeat_region       811. .1058
                    /rpt_family="AluJb"
                    /rpt_type=dispersed
variation           1184
                    /frequency="0.01"
                    /replace="G"
variation           1631. .1643
                    /frequency="0.64"
                    /replace=""
misc_feature        1645. .2063
                    /note="Region not scanned for variation"
gene                1807. .21315
                    /gene="F2R"
mRNA                join(1807. .2237,18156. .21315)
                    /gene="F2R"
                    /product="coagulation factor II (thrombin) receptor"
variation           2081
                    /gene="F2R"
                    /frequency="0.01"
                    /replace="A"
CDS                 join(2150. .2237,18156. .19345)
                    /gene="F2R"
                    /codon_start=1
                    /product="coagulation factor II (thrombin) receptor"
                    /protein_id="AAK69768.1"
                    /db_xref="GI:14583135"
                    /translation="MGPRRLLLVAACFSLCGPLLSARTRARRPESKATNATLDPRSFL
                    LRNPNDKYEPFWEDEEKNESGLTEYRLVSINKSSPLQKQLPAFISEDASGYLTSSWLT
                    LFVPSVYTGVFVVSLPLNIMAIVVFILKMKVKKPAVVYMLHLATADVLFVSVLPFKIS
                    YYFSGSDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTLGR
                    ASFTCLAIWALAIAGVVPLLLKEQTIQVPGLNITTCHDVLNETLLEGYYAYYFSAFSA
                    VFFFVPLIISTVCYVSIIRCLSSSAVANRSKKSRALFLSAAVFCIFIICFGPTNVLLI
                    AHYSFLSHTSTTEAAYFAYLLCVCVSSISCCIDPLIYYYASSECQRYVYSILCCKESS
                    DPSSYNSSGQLMASKMDTCSSNLNNSIYKKLLT"
variation           2372
                    /gene="F2R"
                    /frequency="0.05"
                    /replace="A"
variation           2502
                    /gene="F2R"
                    /frequency="0.01"
                    /replace="A"
variation           2762
                    /gene="F2R"
                    /frequency="0.06"
                    /replace="T"
variation           2860
                    /gene="F2R"
                    /frequency="0.40"
                    /replace="A"
variation           2930
                    /gene="F2R"
                    /frequency="0.22"
                    /replace="C"
variation           2940
                    /gene="F2R"
                    /frequency="0.01"
                    /replace="G"
variation           2952
                    /gene="F2R"
                    /frequency="0.01"
                    /replace="T"
variation           3252
                    /gene="F2R"
                    /frequency="0.01"
                    /replace="C"
variation           3481
                    /gene="F2R"
                    /frequency="0.39"
                    /replace="C"
variation           3525
                    /gene="F2R"
                    /frequency="0.10"
                    /replace="T"
repeat_region       3941. .4047
                    /rpt_family="MIR"
                    /rpt_type=dispersed
variation           3948
                    /gene="F2R"
                    /frequency="0.21"
                    /replace="G"
variation           4017
                    /gene="F2R"
                    /frequency="0.24"
                    /replace="C"
variation           4038
                    /gene="F2R"
                    /frequency="0.03"
                    /replace="A"
variation           4087
                    /gene="F2R"
                    /frequency="0.02"
                    /replace="T"
variation           4107
                    /gene="F2R"
                    /frequency="0.03"
                    /replace="T"
variation           4336
                    /gene="F2R"
                    /frequency="0.41"
                    /replace="A"
variation           4393
                    /gene="F2R"
                    /frequency="0.03"
                    /replace="A"
variation           4617
                    /gene="F2R"
                    /frequency="0.07"
                    /replace="T"
variation           4925
                    /gene="F2R"
                    /frequency="0.01"
                    /replace="G"
repeat_region       4971. .5125
                    /rpt_family="MIR"
                    /rpt_type=dispersed
variation           5072
                    /gene="F2R"
                    /frequency="0.01"
                    /replace="C"
variation           5128
                    /gene="F2R"
                    /frequency="0.22"
                    /replace="G"
repeat_region       5150. .5441
                    /rpt_family="AluSx"
                    /rpt_type=dispersed
variation           5159
                    /gene="F2R"
                    /frequency="0.02"
                    /replace="T"
variation           5210
                    /gene="F2R"
```

```
                     /frequency="0.01"
                     /replace="C"
     variation       5276
                     /gene="F2R"
                     /frequency="0.03"
                     /replace="A"
     variation       5428
                     /gene="F2R"
                     /frequency="0.01"
                     /replace="A"
     variation       5472
                     /gene="F2R"
                     /frequency="0.40"
                     /replace="A"
     variation       5480
                     /gene="F2R"
                     /frequency="0.01"
                     /replace="A"
     variation       5515
                     /gene="F2R"
                     /frequency="0.01"
                     /replace="A"
```

Query Match 55.1%; Score 302; DB 9; Length 24771;
Best Local Similarity 100.0%; Pred. No. 3.3e-57;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     247 CAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCA 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18153 CAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCA 18212

Qy     307 ATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAAT 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18213 ATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAAT 18272

Qy     367 ACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCT 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18273 ACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCT 18332

Qy     427 CAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGT 486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18333 CAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGT 18392

Qy     487 ACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCC 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18393 ACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCC 18452

Qy     547 TG 548
           ||
Db   18453 TG 18454
```

RESULT 11
AC020927

```
LOCUS       AC020927               110965 bp    DNA     linear   HTG 19-APR-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2153E7, WORKING DRAFT SEQUENCE,
            4 ordered pieces.
ACCESSION   AC020927
VERSION     AC020927.4  GI:7711631
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 110965)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 110965)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 6, 2000 this sequence version replaced gi:7417650.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 686404
            Center clone name: CITB-H1_2153E7
            -------
            Summary Statistics
            Consensus quality: 106980 bases at least Q40
            Consensus quality: 109994 bases at least Q30
            Consensus quality: 110455 bases at least Q20
            Estimated insert size: 110000; pulse field gel estimation
            Estimated insert size: 110815; sum-of-contigs estimation
            Quality coverage:  7.03 in Q20 bases; pulse field gel estimation
            Quality coverage:  6.98 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     8728: contig of 8728 bp in length
            *     8729     8828: gap of unknown length
            *     8829    11329: contig of 2501 bp in length
            *    11330    11429: gap of unknown length
            *    11430    40589: contig of 29160 bp in length
            *    40590    40689: gap of unknown length
            *    40690   110965: contig of 70276 bp in length.
FEATURES            Location/Qualifiers
     source         1. .110965
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="CTD-2153E7"
                    /clone_lib="CalTech human BAC library D"
BASE COUNT    32294 a  23208 c  24149 g  31014 t    300 others
ORIGIN
```

Query Match 55.1%; Score 302; DB 2; Length 110965;
Best Local Similarity 100.0%; Pred. No. 3.4e-57;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     247 CAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCA 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   44895 CAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCA 44954

Qy     307 ATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAAT 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   44955 ATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAAT 45014

Qy     367 ACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCT 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   45015 ACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCT 45074

Qy     427 CAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGT 486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   45075 CAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGT 45134

Qy     487 ACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCC 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   45135 ACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCC 45194

Qy     547 TG 548
           ||
Db   45195 TG 45196
```

```
RESULT 12
AC025188
LOCUS       AC025188              143335 bp    DNA     linear   PRI 12-DEC-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2384B11, complete sequence.
ACCESSION   AC025188
VERSION     AC025188.8  GI:17530707
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143335)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143335)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 143335)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 143335)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 12, 2001 this sequence version replaced gi:16328263.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .143335
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2384B11"
BASE COUNT    42376 a  29536 c  30902 g  40521 t
ORIGIN

  Query Match           55.1%;  Score 302;  DB 9;  Length 143335;
  Best Local Similarity 100.0%;  Pred. No. 3.4e-57;
  Matches  302;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     247 CAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCA 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73894 CAGAATCAAAAGCAACAAATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCA 73953

Qy     307 ATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAAT 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73954 ATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAAT 74013

Qy     367 ACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCT 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74014 ACAGATTAGTCTCCATCAATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCT 74073

Qy     427 CAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGT 486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74074 CAGAAGATGCCTCCGGATATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGT 74133

Qy     487 ACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCC 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74134 ACACCGGAGTGTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCC 74193

Qy     547 TG 548
           ||
Db   74194 TG 74195
```

```
RESULT 13
I49726
LOCUS       I49726                 1764 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 12 from patent US 5641650.
ACCESSION   I49726
VERSION     I49726.1  GI:2471946
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1764)
  AUTHORS   Turner,G.J. and Betlach,M.C.
  TITLE     Expression of heterologous polypeptides in halobacteria
  JOURNAL   Patent: US 5641650-A 12 24-JUN-1997;
FEATURES             Location/Qualifiers
     source          1. .1764
                     /organism="unknown"
BASE COUNT      382 a    492 c    421 g    469 t
ORIGIN

  Query Match           51.5%;  Score 282.4;  DB 6;  Length 1764;
  Best Local Similarity 99.6%;  Pred. No. 7.3e-53;
  Matches  283;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     265 ATGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCAT 324
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 ACGCCACCTTAGATCCCCGGTCATTTCTTCTCAGGAACCCCAATGATAAATATGAACCAT 520

Qy     325 TTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCA 384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 TTTGGGAGGATGAGGAGAAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCA 580

Qy     385 ATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGAT 444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 ATAAAAGCAGTCCTCTTCAAAAACAACTTCCTGCATTCATCTCAGAAGATGCCTCCGGAT 640

Qy     445 ATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAG 504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 ATTTGACCAGCTCCTGGCTGACACTCTTTGTCCCATCTGTGTACACCGGAGTGTTTGTAG 700

Qy     505 TCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 548
           |||||||||||||||||||||||||||||||||||||||||||||
Db     701 TCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTCATCCTG 744
```

```
RESULT 14
AC008933
LOCUS       AC008933               88881 bp    DNA     linear   HTG 19-APR-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2297A3, WORKING DRAFT SEQUENCE,
            10 ordered pieces.
ACCESSION   AC008933
VERSION     AC008933.6  GI:13676982
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 88881)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 88881)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 19, 2001 this sequence version replaced gi:9964732.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
```

```
-------
Project Information
Center Project Name: 741600
Center clone name: CITB-H1_2297A3
-------
Summary Statistics
Consensus quality: 79634 bases at least Q40
Consensus quality: 85629 bases at least Q30
Consensus quality: 87035 bases at least Q20
Estimated insert size: 87000; pulse field gel estimation
Estimated insert size: 88431; sum-of-contigs estimation
Quality coverage:  7.1 in Q20 bases; pulse field gel estimation
Quality coverage:  6.99 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*        1     7385: contig of 7385 bp in length
*     7386     7485: gap of unknown length
*     7486     9306: contig of 1821 bp in length
*     9307     9406: gap of unknown length
*     9407    11307: contig of 1901 bp in length
*    11308    11407: gap of unknown length
*    11408    53214: contig of 41807 bp in length
*    53215    53314: gap of unknown length
*    53315    61631: contig of 8317 bp in length
*    61632    61731: gap of unknown length
*    61732    69687: contig of 7956 bp in length
*    69688    69787: gap of unknown length
*    69788    72346: contig of 2559 bp in length
*    72347    72446: gap of unknown length
*    72447    77007: contig of 4561 bp in length
*    77008    77107: gap of unknown length
*    77108    85406: contig of 8299 bp in length
*    85407    85506: gap of unknown length
*    85507    88881: contig of 3375 bp in length.
FEATURES             Location/Qualifiers
     source          1. .88881
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2297A3"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    26315 a  17238 c  18000 g  26421 t    907 others
ORIGIN

  Query Match             45.4%;  Score 249;  DB 2;  Length 88881;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-45;
  Matches  249;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87660 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 87719

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87720 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 87779

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87780 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 87839

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87840 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 87899

Qy     241 GCAGGCCAG 249
           |||||||||
```

```
Db  87900 GCAGGCCAG 87908


RESULT 15
HSU63331
LOCUS       HSU63331                2895 bp    DNA     linear   PRI 18-OCT-1996
DEFINITION  Human thrombin receptor (HTR) gene, partial cds.
ACCESSION   U63331
VERSION     U63331.1  GI:1480846
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2895)
  AUTHORS   Li,F., Baykal,D., Horaist,C., Yan,C.N., Carr,B.N., Rao,G.N. and
            Runge,M.S.
  TITLE     Cloning and identification of regulatory sequences of the human
            thrombin receptor gene
  JOURNAL   J. Biol. Chem. 271 (42), 26320-26328 (1996)
  MEDLINE   96421637
   PUBMED   8824285
REFERENCE   2  (bases 1 to 2895)
  AUTHORS   Li,F., Baykal,D., Hazer,D.J., Horaist,C.K. and Runge,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-1996) Sealy Center for Molecular Cardiology,
            University of Texas Medical Branch, 301 University Boulevard, Rt.
            1064, Galveston, TX 77555-1064, USA
FEATURES             Location/Qualifiers
     source          1. .2895
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="genomic DNA bacteriophage P1 (Genome Systems,
                     Inc.)"
     gene            2811. .2895
                     /gene="HTR"
     CDS             2811. .>2895
                     /gene="HTR"
                     /codon_start=1
                     /product="thrombin receptor"
                     /protein_id="AAC50740.1"
                     /db_xref="GI:1480847"
                     /translation="MGPRRLLLVAACFSLCGPLLSARTRARR"
BASE COUNT      723 a    724 c    853 g    595 t
ORIGIN

  Query Match             44.9%;  Score 246;  DB 9;  Length 2895;
  Best Local Similarity   100.0%;  Pred. No. 1e-44;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2650 CGCCGAGGGTCGCTTGGACCCTGATCTTACCCGTGGGCACCCTGCGCTCTGCCTGCCGCG 2709

Qy      61 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2710 AAGACCGGCTCCCCGACCCGCAGAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 2769

Qy     121 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2770 GGGGCAGCCTCCCGGAGCAGCGCCGCGCAGAGCCCGGGACAATGGGGCCGCGGCGGCTGC 2829

Qy     181 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2830 TGCTGGTGGCCGCCTGCTTCAGTCTGTGCGGCCCGCTGTTGTCTGCCCGCACCCGGGCCC 2889

Qy     241 GCAGGC 246
           ||||||
Db   2890 GCAGGC 2895


Search completed: December  9, 2002, 05:12:19
Job time : 3343 secs
```